Physics-Informed Data-Driven Models for Ship Response Prediction using Global Wave Data

by

Matthew L. Schirmann

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Naval Architecture and Marine Engineering)
in The University of Michigan
2021

Doctoral Committee:

Associate Professor Matthew D. Collette, Co-Chair
Dr. James W. Gose, Co-Chair
Associate Professor Branko Kerkez
Professor Armin W. Troesch

Matthew L. Schirmann

mschirma@umich.edu

ORCID iD: 0000-0003-0790-6238

For my family, and especially my endlessly supportive wife, Rachel.

ACKNOWLEDGEMENTS

I would like to thank the Office of Naval Research, and specifically, Dr. Woei-Min Lin, ONR Code 331, for supporting this research under contract N00014-17-1-2982. I would also like to thank my research advisors and dissertation committee co-chairs Professor Matthew Collette and Dr. James Gose for providing me with this opportunity and invaluable guidance along the way. I owe both of them a great deal for shaping me into the student, researcher, engineer, and professional I am today. I would also like to thank my dissertation committee members Professor Branko Kerkez and Professor Armin Troesch. Professor Troesch deserves extra commendation for being an outstanding resource and teacher throughout my undergraduate and graduate studies. His willingness to go the extra mile and help any "interested student" whenever requested is extremely admirable.

I would like to acknowledge the National Oceanic and Atmospheric Administration (NOAA), as well as the European Union's Copernicus Marine Environment Monitoring Service (CMEMS) and Copernicus Climate Change Service, for providing the invaluable wave data resources used in this research. I would also like to thank Woods Hole Oceanographic Institution, the Scripps Institution of Oceanography, and the Rolling Deck to Repository team for providing a wealth of research vessel data used in this work. Additional thanks are owed to Kongsberg for their technical support regarding onboard measurements. This research was supported in part through computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

I would like to thank colleagues from the VALID II/III joint industry project for sharing ideas and providing valuable data. I would also like to thank Sam Edwards and Claire Wincott for their friendship and willingness to listen to research (or non-research) ideas at any time. I couldn't have asked for better labmates and friends.

Finally, I'd like to thank my family for their unwavering support and encouragement. Mom and Dad, I know that everything you did for us was to make achievements like this possible, and I can't say thank you enough. Rachel, I couldn't ask for a better partner in life, and I appreciate everything you've done to support me over the last four years. I love you all very much.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF ABBREVIATIONS

ARD Automatic Relevance Determination

CDS Climate Data Store (of the Copernicus Climate Change Service)

CMEMS Copernicus Marine Environment Monitoring Service

DOF(s) Degree(s) of Freedom

GPR Gaussian Process Regression

LS Least-Squares

ML Machine Learning

NDBC National Data Buoy Center

NOAA National Oceanic and Atmospheric Administration

NN Neural Network

NWW3 NOAA WAVEWATCH III

PBMP Physics-Based Model Prediction

RAO Response Amplitude Operator

RBF Radial Basis Function

RR Ridge Regression

SIO Scripps Institution of Oceanography

SLNN Shared-Layer Neural Network

VBM Vertical Bending Moment

WHOI Woods Hole Oceanographic Institution

ABSTRACT

Time-and-place specific wave model data, traditional ship response prediction tools, and onboard measurements present an opportunity to train data-driven models for improved motion and structural response predictions. In a real-world implementation, these predictive models would furnish vessel owners and operators with information to support underway, maintenance, and deployment decisions. As demonstrated in this work using over 16,000 30-minute windows from two operational sister ships, although data-driven models are powerful enough to outperform traditional seakeeping and structural response predictions in many instances, retaining physics-based information in these models is paramount to consistent performance improvement. To further incorporate physics-based information, a novel, neural network structure was developed that leveraged shared layers to enforce consistent physics in multidirectional wave data for response prediction. Unlike traditional NN structures trained in this work, this shared-layer approach allowed variation in the number of wave directions considered between samples of multidirectional wave data, as demonstrated using three unique wave data sources with 3 to 24 discrete spectral directions defined for a given time and location. The developed structure also allowed robust wave data interpolation in space and time.

In addition to constructing physics-informed data-driven models, a study was conducted regarding the significance of wave data source selection for response prediction and fatigue damage tracking. This study considered a full year of data from several wave data sources at three separate locations to quantify the resultant differences in response predictions and accumulated fatigue damage. Additionally, a novel approach

was developed for RAO corrections using full- or model-scale seakeeping data and Gaussian process regression. Like other models in this work, the developed approach leveraged physics to enforce that the RAO correction reverted to physics-based RAOs in regions of the input space far from train samples. A final demonstration combining this RAO correction approach with the shared-layer neural network was performed using the full dataset from the two sister vessels. The results of this work demonstrate significant promise for real-world implementation of data-driven response prediction models, especially those that retain physics-based information.

CHAPTER I

Introduction

Time-and-place specific wave forecast and hindcast data covering a majority of the world's oceans are now widely available, characterizing a vessel's operational environment days in advance and throughout its life cycle. When investigated with traditional, frequency-domain motion and structural response analyses, these wave data hold a wealth of information regarding the ship's past, present, and future experiences. However, the insights available through these global wave data extend far beyond the output of traditional response prediction tools. Many modern ships carry continuously-recording motion measurement devices, often implemented for use by other instruments critical to standard vessel operation. These onboard motion measurements or structural response data, when available, provide an exciting opportunity to extend wave-data-based response prediction capabilities through data-driven models with embedded machine learning (ML) algorithms, which have proven extremely powerful for learning complex relationships in both classification and regression tasks. As discussed in the context of vessel response forecasting and structural fatigue damage tracking, these improved response predictions would provide vessel owner's and operators with valuable guidance for underway, deployment, and maintenance decisions. Using over 16,000 30-minute windows from two operational sister ships, this body of work demonstrates the fusion of global wave data, in-service

measurements, and design-stage analyses through data-driven approaches, yielding significant improvement in predictive capabilities.

In pursuit of data-driven models for improved response prediction, the benefits of retaining physics-based model predictions (PBMPs) from linear, frequency-domain theory were explicitly demonstrated, combating the criticism that many data-driven models abandon physics entirely. These PBMPs were generated using wave model data at the ship's known locations and corresponding speed and heading information from the ship. These PBMPs were then provided alongside the speed, heading, and wave data as input to linear and neural network (NN) response prediction models. In addition to demonstrating the importance of retaining physics-based information in data-driven models, this work also developed a novel, physics-informed structure for use with multidirectional wave data called a Shared-Layer Neural Network (SLNN). This structure significantly improved upon the underlying PBMPs and was flexible for use with multidirectional wave data sources that had a varying number of directional wave partitions, which the traditional NN structure was unsuited for. SLNN development also yielded a unique interpolation approach for gridded wave data in space and time.

Finally, this work examined the use of model- and full-scale motions data to directly target seakeeping model inadequacies through correction of response amplitude operators (RAOs) via Gaussian process regression (GPR). A rigorous wave data filtering approach was developed to attribute encountered response spectral energy from full-scale, onboard measurements to specific wave frequencies, even in following and stern-quartering seas, such that RAOs could be calculated for GPR model training. A two-stage model demonstration, combining this GPR RAO correction approach with the SLNN structure, was performed to investigate potential for further improvement in vessel motion prediction capabilities.

The contributions discussed herein have been presented across several conferences

(*Schirmann et al.* (2018a), *Schirmann et al.* (2018b), *Schirmann et al.* (2019a), *Schirmann et al.* (2019b), *Schirmann et al.* (2020b)), and two journal submissions (*Schirmann et al.* (2020a) and *Schirmann et al.* (2021)).

1.1 Research Objectives

The primary research objectives addressed in this dissertation are as follows:

1. Explore the use of time-and-place specific wave data for response prediction and fatigue damage tracking.

2. Develop data-driven approaches for frequency-domain motion and structural response prediction that improve upon traditional, physics-based methods.

3. Evaluate the importance of retaining physics-based information in data-driven response prediction models.

4. Identify approaches for updating underlying, physics-based response prediction models with full- and model-scale data.

1.2 Viewpoint as a Digital Twin Application

A digital twin is a dynamic virtual representation of a physical system for the purposes of operational insight and life-cycle management. As the physical system and its environment change continuously, this technology seeks to link, compare, and fuse physical system measurements and observations with numerical models such that the models accurately reflect the system's current state. The digital twin concept has been developed and applied in several fields including the aerospace industry for structural integrity management and by General Electric for machinery performance monitoring, as discussed by *Glaessgen and Stargel* (2012) and *Egan* (2015), respectively. Although "digital twin" is a frequently-abused buzzword with a wide

variety of interpretations, the coupling of global wave data, physics-based response prediction models, onboard measurements, and data-driven models with embedded machine learning (ML) algorithms is a prime example of digital twin technology in the marine context.

As alluded to previously, numerical models for surface ship performance prediction and condition assessment, which are used extensively throughout the design process, are often neglected after the ship has begun operation because too many variables affect the physical system and separate it from the idealized cases investigated during the design phase. As demonstrated in this work, the ability to link the ship to these virtual models would permit the inherent capabilities of the models to remain applicable throughout the life of the ship, and thus allow operators to make more informed decisions underway and owners to schedule fleet-wide deployments and maintenance operations more effectively. For instance, motion and structural response forecasts and hindcasts, rather than wave data alone, would provide owners and operators with valuable information to support deployment and underway decisions. These benefits could be significantly enhanced using onboard motion and structural response measurements to train data-driven models for improved predictive performance. Even without onboard instrumentation, cumulative fatigue damage estimation and tracking along a vessel's known routes using these wave data would further support fleet-wide maintenance and deployment decisions. If all ships in a fleet were constructed at a similar time, but the ships are exposed to different environments throughout the course of their operation, some may accumulate fatigue damage more rapidly than others due to differences in operational profiles and encountered sea states. If a data stream between the ship and these virtual model were opened to quantify the vessel's fatigue damage accumulation based on encountered seas over the course of its missions, more informed maintenance decisions could be made. Cumulative fatigue damage between ships with similar capabilities could also

be compared to make educated deployment decisions such that the vessels maintain fatigue damage equality.

To put this application in perspective, a conceptual digital twin schematic for a surface platform is included in Figure 1.1.

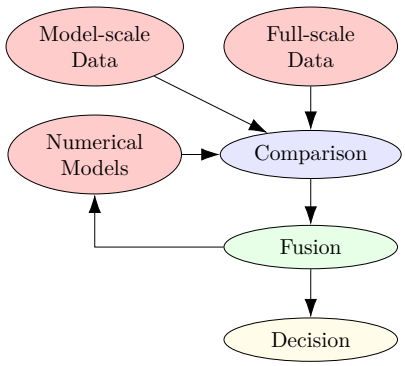

Figure 1.1: A simplified conceptual schematic of a surface platform digital twin.

As shown in Figure 1.1, numerical models regarding the vessel's state (e.g., environment, response, structural condition, and machinery performance) are compared to full-scale data from the vessel, which may include both sensor data and operator observations. Model-scale experimental data can also be used to address discrepancies between numerical models and real-world measurements in the comparison step. The numerical models and full-scale observations are then fused appropriately based on comparisons and applicable model-scale data to predict the future status of the vessel. The results of the fusion process may also be used to update the underlying numerical models for future predictions. The fusion stage is required in a reflective, or updating twin framework. A more preliminary twin that does not consider real-world observations as feedback for model updating is referred to as "forward-only" in this work.

The contributions from this work primarily address the comparison and fusion stages of the conceptual digital twin in Figure 1.1. In this application, the numerical models were comprised of wave hindcast and forecast models, motion and structural response prediction analyses, and structural health monitoring in the form of accu-

mulated fatigue damage tracking. Full-scale data were considered for training the physics-informed, data-driven response prediction models, which reside in the fusion stage of Figure 1.1. The GPR RAO correction models trained using full- and model-scale data also reside in the fusion stage; however, the resultant GPR-corrected RAOs represent an update that was passed back to the underlying numerical models. In a real-world application, the response predictions and accumulated fatigue damage produced by the fusion stage would then be passed to the decision stage where owners and operators could make informed choices regarding the vessel's current mission or deployment and maintenance schedules.

1.3 Overview of Related Work

This section gives a general overview of related work most relevant to the initial twin framework and wave data source selection studies of Chapters II and III, with literature reviews more specific to the data-driven models included in subsequent chapters. *Thompson* (2018) introduced and evaluated the concept of virtual hull monitoring using a vessel's known location and measured or numerically modeled wave data. The investigation used data from a naval vessel trial and included a comparison of wave hindcast data to those collected via a deployed wave buoy, as well as a comparison between stresses derived from strain gauge measurements and those predicted using a spectral fatigue analysis and the hindcast data. *Hulkkonen et al.* (2019) also discussed the use of wave model data and a vessel's known locations to track the fatigue damage accumulated by the vessel, as well as the uncertainties associated with this method of virtual monitoring (e.g. vessel loading conditions, corrosion, etc.). Similar studies, such as *Mondoro et al.* (2016) and *Zhu and Collette* (2017) have considered means of structural health monitoring using limited onboard response measurements to update a vessel's operational profile and expected responses in unseen operating conditions.

Magoga et al. (2019) proposed a fatigue life updating method based on in-service measurements and maintenance data, and *Nielsen et al.* (2011) explored the implementation of a real-time onboard decision support system that considers fatigue damage accumulation rates. A more recent study conducted by *Thompson* (2020) further investigated the accuracy of virtual hull monitoring via wave hindcast data using strain gauge measurements from a naval vessel. The study compared structural responses calculated using wave data from different hindcast data sources. While the results of *Thompson* (2020) were very promising for this method of structural health monitoring and showed agreement between stress spectra generated using different wave hindcast sources, it was noted that the data considered were limited to a two-week cruise with a relatively narrow set of operating conditions. In contrast, the wave data study of Chapter III considers a full year of wave data, response predictions, and accumulated fatigue damage at multiple locations.

The aforementioned studies all address structural health monitoring using wave data, in-service measurements, or both. Furthermore, many of these studies can be considered digital twin applications. The novelty of this thesis lies in its fusion of wave model data, traditional response prediction models, and data-driven approaches for improved response prediction capabilities. These data-driven approaches are necessary to shift this digital twin application from "forward-only" to reflective.

1.4 Overview of Chapters

In this work, an initial twin framework was developed that was capable of linear, frequency-domain seakeeping and structural response predictions and fatigue damage tracking using wave forecast or hindcast data, which is discussed in Chapter II. This initial twin was "forward-only," as defined in Section 1.2, because it did not use observed vessel behavior as feedback to update future predictions, which would be required in a reflective, or updating, twin framework. This forward-only, initial

twin framework served as the physics-based model for generation of PBMPs used in subsequent data-driven response prediction models, which represent reflective twin capabilities.

A demonstration of the initial twin framework is also presented in Chapter II that compared four voyages of a naval combatant in the Pacific, which highlighted this technology's potential for fatigue damage balancing between ships in a fleet. This demonstration also spawned a critical question, addressed in Chapter III, for reliable response prediction - What is the effect of wave data source selection on the initial twin's response predictions and accumulated fatigue damage estimates? To address this inquiry, multiple locations in the Pacific Ocean were selected, each with relevant wave data available from three different sources, and vessel response simulations were run for the entirety of 2017 using each source.

After demonstration of the initial twin and the wave data source selection study, full-scale motion data were acquired from a research vessel to begin exploration of data-driven prediction models. As discussed in Chapter IV, two preliminary approaches were tested with unidirectional wave data: a linear least-squares (LS) model and an NN. Using full-scale measurements to train the correction models, the success of these relatively simplistic approaches showed promise for incorporation of more intensive statistical and machine learning approaches. Chapter IV also discusses a successful experiment using onboard vertical bending moment (VBM) measurements to train and test LS and NN models. Given these successes, improved linear and NN models were then trained and tested using more detailed multidirectional wave data and a dataset comprised of over 16,000 30-minute windows from two operational sister ships, as outlined in Chapter V. The consideration of data from two similar ships allowed investigation of potential fleet-wide applicability for these data-driven models. The large quantity of available data also allowed formal investigation into the benefits of retaining physics-based information in the data-driven models via PBMP

input variables.

Chapter VI describes the physics-informed SLNN structure developed in this work and demonstrates its robustness and versatility using three unique wave data sources and the full dataset from Chapter V. Chapter VII discusses GPR RAO correction using model-scale data, and Chapter VIII discusses the rigorous wave data filtering algorithm developed for GPR RAO correction using full-scale data. Chapter VIII also discusses a final, two-stage model demonstration in which the GPR-corrected RAOs were used in the initial twin framework to generate PBMPs (stage one), and these PBMPs were used to train and test the SLNN (stage two). The results of both stages were then directly compared to their counterparts from Chapter VI to investigate the potential benefits of combining these approaches for improved response prediction.

CHAPTER II

Initial Twin Framework and Demonstration

2.1 Initial Twin Framework

An initial twin framework was developed for frequency-domain vessel motion and structural response predictions using wave forecast or hindcast data. Although considered "forward-only" as defined in Chapter I, this framework was the foundation for PBMP generation in all subsequent chapters of this dissertation. Beyond the wave data source selection study of Chapter III, these chapters all addressed reflective twin capabilities using PBMPs for performance comparisons, as input to the data-driven models, or both. Additionally, this initial twin framework was extended to estimate and track fatigue damage in time using predicted vertical bending moments (VBMs) and Miner's sums, as demonstrated in this chapter and Chapter III.

The information necessary for the framework to run included a list of route waypoints (defined by a date/time, location, speed, and heading), a wave data file containing time-and-place specific wave data relevant to the route waypoints, and a collection of response amplitude operators (RAOs) for different speeds and headings relative to the incoming waves. The framework was modular because it allowed the wave data source and numerical model used for RAO generation to be changed. A demonstration of this framework was performed using the *DTMB 5415*, a destroyer-sized naval combatant described by *Stern et al.* (2011), on four concurrent voyages in

the Pacific, which was presented in *Schirmann et al.* (2018b). Table 2.1 provides select characteristics of the *DTMB 5415* that were used in hydrodynamic modeling. While the results of this demonstration are presented in Section 2.2, the discussion in this section uses the details of the initial twin demonstration to examine the framework's calculations. The initial twin framework was also discussed extensively in *Schirmann et al.* (2018a), *Schirmann et al.* (2019a), *Schirmann et al.* (2019b), *Schirmann et al.* (2020a), *Schirmann et al.* (2020b), and *Schirmann et al.* (2021).

Table 2.1: Select characteristics of the *DTMB 5415*.

Length Between Perpendiculars	142.0 m
Beam at Waterline	19.1 m
Draft	6.1 m
VCG (Above Waterline)	1.0 m
Block Coefficient	0.507
Displacement	8,488 LT

2.1.1 Input Wave Data

Time-and-place specific wave data are commonly defined using multiple wave partitions, each representing a fraction of the total wave energy with an associated direction. Each of these directional wave partitions are typically defined by characteristic parameters (i.e. a significant wave height, H_s, and period, T). Given these characteristic parameters for each directional wave partition, the initial twin framework could construct an idealized wave spectrum (e.g. JONSWAP, Bretschneider). Note that the developed framework was also compatible with wave partitions characterized by defined spectral shapes (e.g., wave buoy data, ship wave radar measurements), as demonstrated in Chapters III, VI, and VIII. In Chapters VI and VIII, this allowed response prediction using regular gridded 2D (wave frequency and direction) spectra from a more detailed wave model dataset.

For the initial twin demonstration, wave forecast data generated using the NOAA WAVEWATCH III (NWW3) wave model were used, which were defined for a given

time (in three-hour increments) and location (in 0.5° latitude/longitude increments) by three wave partitions: wind waves (WW), primary swell ($S1$), and secondary swell ($S2$) (*The WAVEWATCH III® Development Group (WW3DG)*, 2016). Each directional partition was defined by the characteristic wave parameters mentioned above, with the period defined as the mean period, $\overline{T}$, in this case. These parameters were used to construct a Bretschneider (ITTC two-parameter) spectrum, $S^{+}_{WP}(\omega_0)$, for each wave partition (WP) using Equations 2.1-2.3, where ω_0 is the wave frequency (*Lloyd*, 1989).

$$S^{+}_{WP}(\omega_0) = \frac{A}{\omega_0^5} exp\left(\frac{-B}{\omega_0^4}\right) \tag{2.1}$$

$$A = 172.75\frac{H_s^2}{\overline{T}^4} \tag{2.2}$$

$$B = \frac{691}{\overline{T}^4} \tag{2.3}$$

The idealized wave spectrum for each of the directionally defined wave partitions was then used to generate the vessel's response to each wave partition independently, as detailed in the following section. Construction of an idealized spectrum for each wave partition was only required in the initial twin demonstration because defined spectral shapes were unavailable.

2.1.2 Response Prediction

The initial twin framework generated frequency-domain motion and structural response predictions using RAOs. For the initial twin demonstration, heave, pitch, and VBM RAOs were calculated using `SHIPMO.BM`, a frequency-domain, 2D strip theory code developed by *Beck and Troesch* (1990). Note that the roll calculations of `SHIPMO.BM` are based on the method of *Himeno* (1981). Also note that the RAO for the j^{th} degree of freedom (DOF) or structural response, $RAO_j(\omega_0)$, was extracted from `SHIPMO.BM` as a function of wave frequency, ω_0, rather than encounter frequency,

ω_e. This choice was made for the twin framework to avoid numerical integration complications that arise in following seas when multiple wave frequencies result in the same encounter frequency, as discussed in *Lloyd* (1989). This consideration is revisited in the following paragraph and Chapters VII and VIII. Also note that $RAO_j(\omega_0)$ was additionally a function of vessel speed and heading relative to the incoming waves. Head seas were defined as 180° and following seas were defined as 0°. This frequency-domain approach relied upon the assumptions that the vessel experienced stationary and ergodic wave conditions, and the vessel's response spectrum was a linear function of the wave spectrum.

When vessels have non-zero forward speed in following or stern-quartering seas, wave frequencies can yield zero or negative encounter frequencies. While reasonable RAOs can be calculated for negative encounter frequencies, 2D strip theory tends to yield non-physical results at encounter frequencies near zero. As such, RAOs were "corrected" at wave frequencies near the wave frequency that yielded an encounter frequency equal to zero, w_{0e0}, for a given speed and relative wave direction. This correction was achieved using cubic spline interpolation (implemented with `interpolate.splrep` of `Python`'s `SciPy` library) as detailed in the following paragraph (*Virtanen et al.*, 2020). Examples of this correction for heave and pitch are demonstrated in Figure 2.1 for the *DTMB 5415*.

All RAOs generated using `SHIPMO.BM` in this dissertation used wave frequency steps of 0.025 rad/s beginning at a minimum value 0.01 rad/s. In this initial twin demonstration, these frequency steps continued to a maximum frequency of 1.21 rad/s, yielding 49 wave frequencies. For all chapters after the present demonstration, the maximum frequency was increased to 4.885 rad/s (196 total wave frequencies). Due to limitations on the number of wave frequencies usable in `SHIPMO.BM` at once, note that the program was run for several frequency ranges that were subsequently pieced together. Through experimentation, it was determined that replacing the

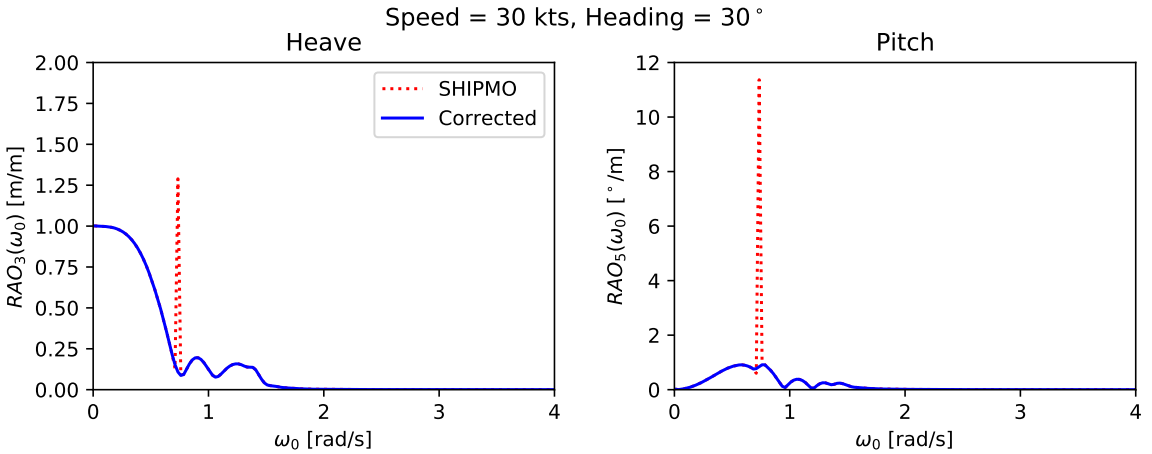

Figure 2.1: Heave and pitch RAOs generated using `SHIPMO.BM` and their cubic spline-corrected counterparts for the *DTMB 5415* at 30 knots and a relative wave heading of 30°.

RAO values of the three discrete wave frequencies nearest w_{0e0} sufficiently removed the non-physical RAO behavior across all speeds and headings. Of course, changing the discrete wave frequency step size may warrant replacement of a different number of RAO values. Given the nine wave frequencies and corresponding RAO values closest to w_{0e0}, the six RAO values not marked for replacement (furthest from w_{0e0}) were used to fit the cubic spline. This fitted cubic spline was then used to replace the three values nearest w_{0e0}. Figure 2.2 gives another example of heave and pitch RAOs before and after correction, which shows close-ups of the nine discrete RAO values at wave frequencies closest to w_{0e0}. Please note that the RAO correction highlighted in Figures 2.1 and 2.2 was not employed until after completion of the initial twin demonstration in Section 2.2; however, this correction was applied for all `SHIPMO.BM` RAOs in subsequent chapters.

Given the vessel's speed and heading at a given location, the one-sided pseudo-response spectrum due to an individual wave partition, WP, in the j^{th} DOF was calculated using Equation 2.4.

$$S^{+}_{j,WP}(\omega_0) = S^{+}_{WP}(\omega_0)|RAO_j(\omega_0)|^2 \tag{2.4}$$

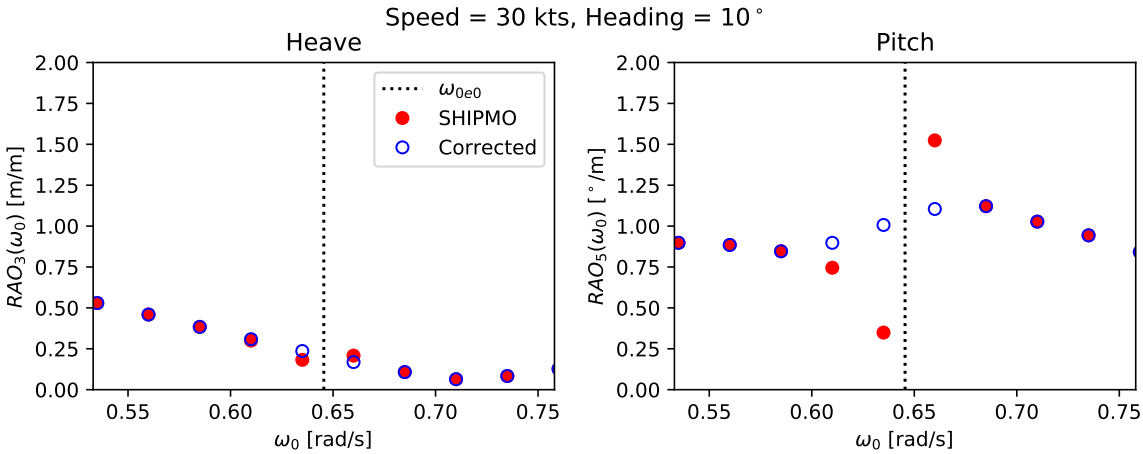

Figure 2.2: Close-ups of cubic spline-corrected `SHIPMO.BM` heave and pitch RAOs near w_{0e0} for the *DTMB 5415* at 30 knots and a relative wave heading of 10°.

The prefix "pseudo-" is used here as it is in *Lloyd* (1989) to denote that the response spectrum is a function of ω_0 rather than ω_e. The k^{th} spectral moment of the vessel's response in the j^{th} DOF due to an individual wave partition, $m_{jk,WP}$, was then calculated through use of trapezoidal integration to solve Equation 2.5.

$$m_{jk,WP} = \int_0^\infty \left(\omega_0 - \frac{\omega_0^2 U}{g}\cos(\mu)\right)^k S^+_{j,WP}(\omega_0)\, d\omega_0 \tag{2.5}$$

In Equation 2.5, U is the speed of the vessel, g is the acceleration due to gravity, and μ is the relative heading of the ship to the incoming wave partition, where μ equal to 180° corresponds to head seas. In Equation 2.5, the first term, raised to the k^{th} power, accounts for the shift between wave frequency and encounter frequency required for calculation of moments where $k > 0$. The total k^{th} spectral moment of the vessel in the j^{th} DOF, $m_{jk,T}$, was calculated through linear superposition of the k^{th} moments resulting from each of the N independent wave partitions as shown in Equation 2.6. This summation is explicitly outlined for the three wave partitions of

the initial twin demonstration in Equation 2.7.

$$m_{jk,T} = \sum_{WP=1}^{N} m_{jk,WP} \tag{2.6}$$

$$m_{jk,T} = m_{jk,WW} + m_{jk,S1} + m_{jk,S2} \tag{2.7}$$

Note that under the assumption of a Gaussian process required for this approach, the zeroth moment of the total response, $m_{j0,T}$, was the variance of the response. Therefore, the standard deviation of the response, $\sigma_{j,T}$, and the response significant amplitude, $\zeta_{j,T}$, were calculated using Equations 2.8 and 2.9, respectively.

$$\sigma_{j,T} = \sqrt{m_{j0,T}} \tag{2.8}$$

$$\zeta_{j,T} = 2\sigma_{j,T} \tag{2.9}$$

For the initial twin demonstration, heave, pitch, and VBM responses were calculated at different route waypoints using this methodology, as examined in Section 2.2.2.

2.1.3 Fatigue Damage Estimation

The initial twin's capabilities were extended to estimate accumulated fatigue damage using the calculated VBM responses. Two fatigue damage estimation methods were employed within the initial twin framework. The first method, used in the initial twin demonstration of Section 2.2, is described in Section 2.1.3.1. The second method, an improvement upon the first because it incorporated spectral bandwidth, is outlined in Section 2.1.3.2. This improved method was employed in the wave data source selection study of Chapter III. In the following sections, VBM responses are indicated with the DOF subscript, j, equal to M.

2.1.3.1 First Fatigue Damage Estimation Method

Frequency-domain fatigue damage estimation relies upon the assumption of narrow-bandedness, which is not guaranteed during real-world operation, especially in the case of multidirectional seas; therefore, it is important to note that the following fatigue damage calculation procedure, which performs the nonlinear calculation of fatigue damage due to each wave partition independently, yields a non-conservative estimate that should be used for comparison purposes only (as in the case of the initial twin demonstration). To estimate the number of loading cycles experienced by the vessel for fatigue damage estimation, it was necessary to calculate the mean peak frequency, $f_{p,WP}$, of the response spectrum due to an individual wave partition as shown in Equation 2.10.

$$f_{p,WP} = \frac{1}{2\pi}\sqrt{\frac{m_{M4,WP}}{m_{M2,WP}}} \tag{2.10}$$

In Equation 2.10, $m_{M2,WP}$ and $m_{M4,WP}$ are the second and fourth moments of the VBM response, respectively, which were calculated using Equation 2.5.

Prior to fatigue damage calculation, the section modulus at a given location of the ship, SM, was used to transform the zeroth moment of the VBM response due to an individual wave partition, $m_{M0,WP}$, into a stress zeroth moment, $m_{S0,WP}$, as shown in Equation 2.11.

$$m_{S0,WP} = \frac{m_{M0,WP}}{SM^2} \tag{2.11}$$

Although SM is a function of longitudinal location and distance from the neutral axis, for the initial twin demonstration, a bottom cross section modulus of 3.77 m^3 was selected from *Ashe et al.* (2009) for the calculation of representative fatigue damage values at all longitudinal stations of the *DTMB 5415*. The nondimensional damage resulting from an individual wave partition, D_{WP}, at a single route waypoint was then calculated using Equation 2.12, which is based on Palmgren-Miner's rule and

was adapted from *Det Norske Veritas AS* (2014).

$$D_{WP} = \frac{T f_{p,WP}}{C}(2\sqrt{2m_{S0,WP}})^m \Gamma(1 + m/2) \tag{2.12}$$

In Equation 2.12, T is the exposure time in seconds, $\Gamma()$ is the gamma function, and m and C are constants corresponding to the log-log S-N curve slope and $y-$intercept, respectively. For the initial twin demonstration, T was equal to 10,800 seconds (3 hours), and values of 3.0 and $10^{12.18}$ were used for m and C, respectively, which were included in *Hughes* (1988) and originally provided in *The Welding Institute* (1976) for class D joints with stress in MPa. The fatigue damage accumulated by the vessel at a given station during a three-hour time step was then non-conservatively estimated through summation of the D_{WP} values from all wave partitions. The accumulated fatigue damage was then tracked over time. In this Palmgren-Miner's rule fatigue damage formulation, the ratios of experienced cycles at each stress level over the number of cycles to failure at that stress level are linearly summed to quantify the fraction of fatigue life consumed. As such, a nondimensional fatigue damage value equal to 1.0, the entirety of the fatigue life, corresponds to structural failure. Note that Palmgren-Miner's rule does not account for load sequence effects (*Det Norske Veritas AS*, 2014).

2.1.3.2 Improved Fatigue Damage Estimation Method

As mentioned previously, after the initial twin demonstration provided in Section 2.2, it was deemed necessary to utilize a fatigue damage estimation approach that addressed the bandwidth of the VBM responses. Following the *Guide for Fatigue Assessment of Offshore Structures* (*American Bureau of Shipping*, 2018) from the American Bureau of Shipping (ABS), the spectral fatigue analysis of *Wirsching and Light* (1980) was implemented, which incorporates a rainflow correction factor based

on bandwidth.

To characterize the number of fatigue cycles experienced during a given three-hour exposure period, it was necessary to calculate the mean zero-up crossing frequency using Equation 2.13.

$$f_z = \frac{1}{2\pi}\sqrt{\frac{m_{M2,T}}{m_{M0,T}}} \tag{2.13}$$

In Equation 2.13, $m_{M0,T}$ and $m_{M2,T}$ are the zeroth and second total VBM spectral moments, respectively, which were calculated using Equations 2.5 and 2.6. Similarly, the bandwidth, ε, of the total VBM response, was calculated using Equation 2.14, which also considers the fourth total VBM spectral moment, $m_{M4,T}$.

$$\varepsilon = \sqrt{1 - \frac{m_{M2,T}^2}{m_{M0,T}\, m_{M4,T}}} \tag{2.14}$$

In Equation 2.15, the standard deviation of the total VBM response, $\sigma_{M,T}$, was transformed into a stress standard deviation, $\sigma_{S,T}$, using the section modulus, SM, at a select location within the cross-section of the ship.

$$\sigma_{S,T} = \frac{\sigma_{M,T}}{SM} \tag{2.15}$$

The damage, D, resulting from all wave partitions at a single route waypoint was then calculated using Equation 2.16 (*American Bureau of Shipping*, 2018; *Wirsching and Light*, 1980), which is also based on Palmgren-Miner's rule. As such, identical to the first fatigue estimation method, a nondimensional fatigue damage value equal to 1.0 corresponds to structural failure.

$$D = \lambda(m,\varepsilon)\frac{T f_z}{C}(2\sqrt{2}\sigma_{s,T})^m \Gamma(1 + m/2) \tag{2.16}$$

In Equation 2.16, $\lambda(m,\varepsilon)$ is Wirsching's rainflow correction factor. As in the first fatigue damage estimation method, T is the exposure time, $\Gamma()$ is the gamma function,

and m and C are constants corresponding to the log-log S-N curve slope and intercept, respectively. Note that the m and C values given in Section 2.1.3.1 for the initial twin demonstration were also used in the wave data source selection study of Chapter III.

Using the bandwidth, ε, and S-N curve constant, m, Wirsching's rainflow correction factor, $\lambda(m, \varepsilon)$, was calculated using Equations 2.17−2.19 (*American Bureau of Shipping*, 2018; *Wirsching and Light*, 1980).

$$\lambda(m, \varepsilon) = a(m) + [1 - a(m)][1 - \varepsilon]^{b(m)} \tag{2.17}$$

$$a(m) = 0.926 - 0.033m \tag{2.18}$$

$$b(m) = 1.587m - 2.323 \tag{2.19}$$

Although the non-conservative first fatigue damage estimation method was reasonable for the comparisons made in the initial twin demonstration of Section 2.2, implementation of this improved fatigue damage estimation procedure allowed more valid comparisons between wave data sources with differing multidirectional wave data formats in Chapter III.

2.2 Initial Twin Demonstration

The following demonstration of the initial twin framework using the *DTMB 5415* was published in *Schirmann et al.* (2018b), and also presented in an ASNE TSS 2018 student poster that was included in *Naval Engineers Journal* (*Schirmann et al.*, 2018a). The four routes compared in the initial twin demonstration are summarized in Table 2.2 and mapped in Figure 2.3, which shows the NWW3 combined significant wave height and vessel locations at a single time step of the simulation. All four voyages began simultaneously at the NWW3 wave data forecast time of 00:00 GMT on December 3, 2017, with responses calculated every three hours corresponding to

the wave data time increments. At each route waypoint, wave data from the nearest NWW3 grid location (every 0.5° latitude/longitude) were used. For each of the four routes, the vessel was assumed to travel at a constant heading and speed of 20 knots for the duration of the voyage.

Table 2.2: Summary of the four routes analyzed for cumulative fatigue damage comparison. Distances are provided in nautical miles (NM) (*SEA-DISTANCES.ORG*, 2018).

Route	Distance [NM]	Time [hours]
Guam-Hawaii	3,333	167
Japan-Hawaii	3,380	169
Seattle-Hawaii	2,403	120
Guam-Japan	1,348	67

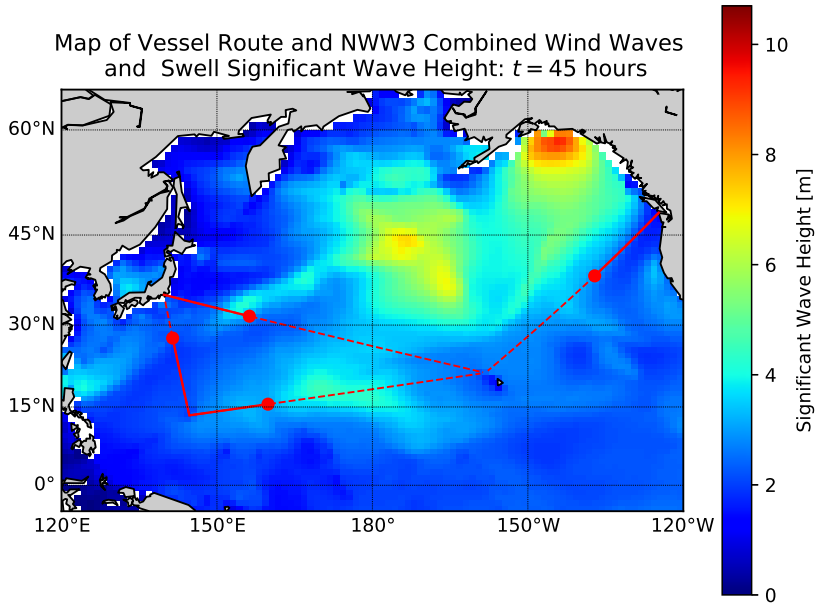

Figure 2.3: A single frame of the animated Pacific Ocean significant wave height map including the four routes analyzed for fatigue damage comparison and the location of each vessel at the current simulation time step.

2.2.1 Fatigue Damage Comparison

The cumulative fatigue damage values at each of the vessel's 21 longitudinal stations for a given route were calculated and normalized by the trip's total time. The resulting cumulative fatigue damage per hour values for Stations 5 through 15 are shown in Figure 2.4. The sets of five stations nearest the bow and stern were excluded from Figure 2.4 because they showed the same fatigue damage relationships between the different routes as the midbody stations with relatively insignificant fatigue damage values for these specific scenarios. Nevertheless, the ability to track fatigue damage at these excluded stations may prove valuable under different circumstances.

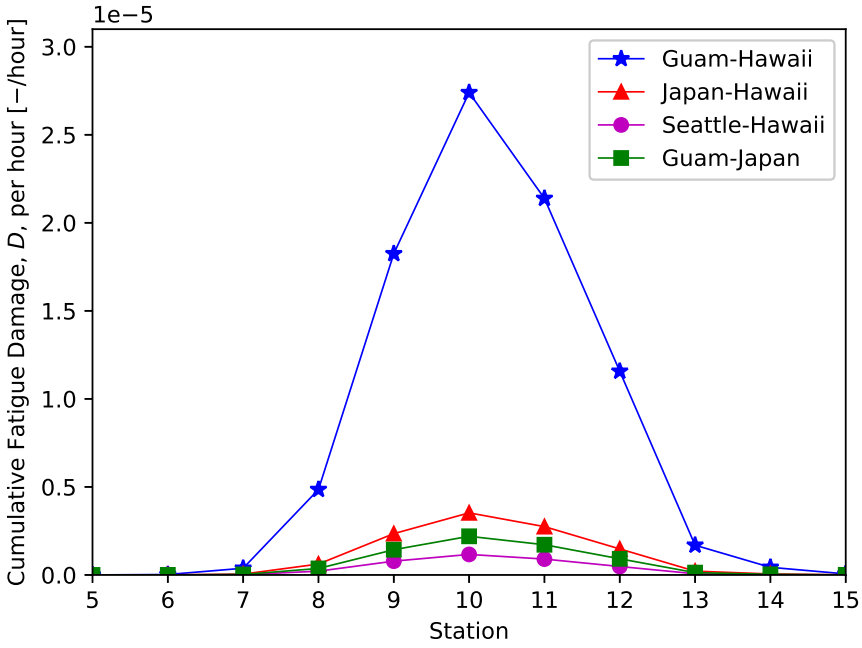

Figure 2.4: Comparison of cumulative fatigue damage, D, per hour at Stations 5 through 15 of the vessel. Station 10 is the midship station, which is typically where the largest VBMs occur.

As shown in Figure 2.4, the most significant fatigue damage value predicted for each of the four routes occurred at Station 10. This result was expected because the largest VBMs in a vessel typically occur amidships. The simulated route from Guam to Hawaii yielded notably greater cumulative fatigue damage values than the other

routes, which is largely attributed to the relatively large head seas encountered along a majority of the journey. These sea conditions and corresponding vessel motions are discussed in Section 2.2.2. As an example of the significance of this fatigue damage at midships, consider an operational time of 5,000 hours per year with an equivalent fatigue damage accumulation rate. The fatigue damage at Station 10 would reach the critical level ($D = 1.0$) in approximately 7 years. Under this same set of assumptions, the route from Japan to Hawaii would not cause critical fatigue damage at midships for approximately 57 years. Although this fatigue life prediction is highly idealized and a naval combatant's missions and encountered sea conditions change frequently, this estimation stresses the importance of the ability to predict and track fatigue damage throughout a vessel's life. Note that this prediction method used real wave data, but did not require data from inspections or installation of monitoring equipment. If results similar to those shown in Figure 2.4 continued for an extended period of time, vessels in the fleet could periodically adjust deployments to balance damage among the fleet. Cumulative fatigue damage tracking would also help to ensure that individual vessels are maintained at appropriate intervals.

2.2.2 Response Prediction

The initial twin framework made time-and-place specific predictions of the significant heave, pitch, and roll amplitudes of a vessel using forecast directional wave partitions. To demonstrate the framework's response prediction capabilities, consider the previous example route from Guam to Hawaii, a journey lasting approximately 167 hours. This route was selected for further discussion because it yielded significantly more cumulative fatigue damage than the other three voyages analyzed in Section 2.2.1. Figure 2.5 shows the vessel's heave, pitch, and roll significant amplitudes calculated at each route waypoint. Figure 2.5 also demonstrates the linear superposition of zeroth moments due to individual wave partitions, as described in

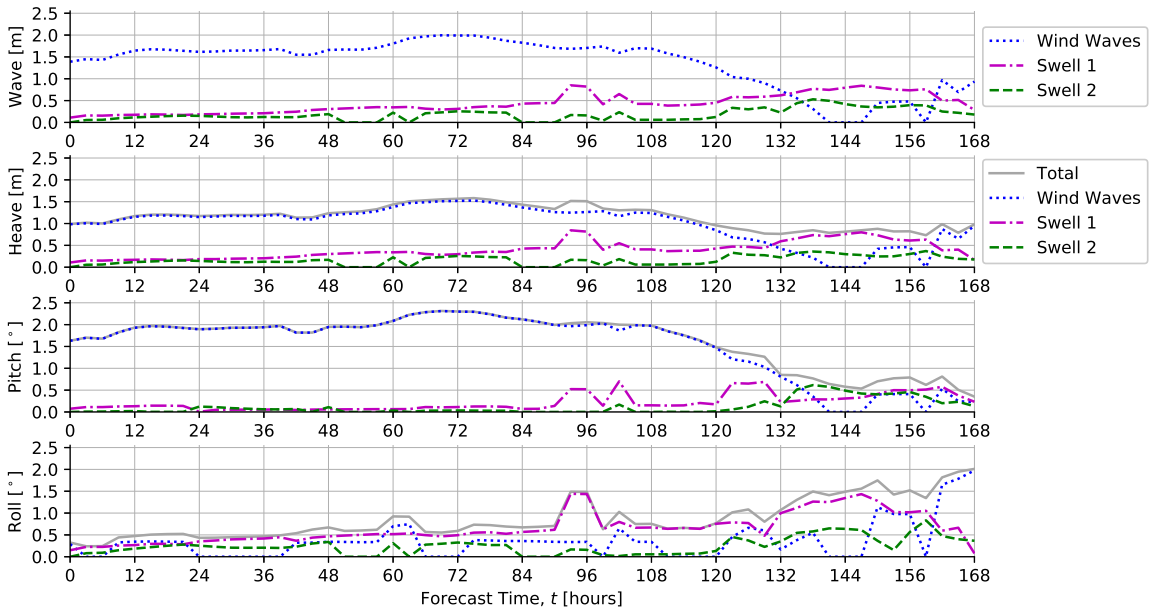

Figure 2.5: Incoming wave partition significant amplitudes and predicted heave, roll, and pitch significant amplitudes of the total responses and responses due to individual wave partitions on a course from Guam to Hawaii.

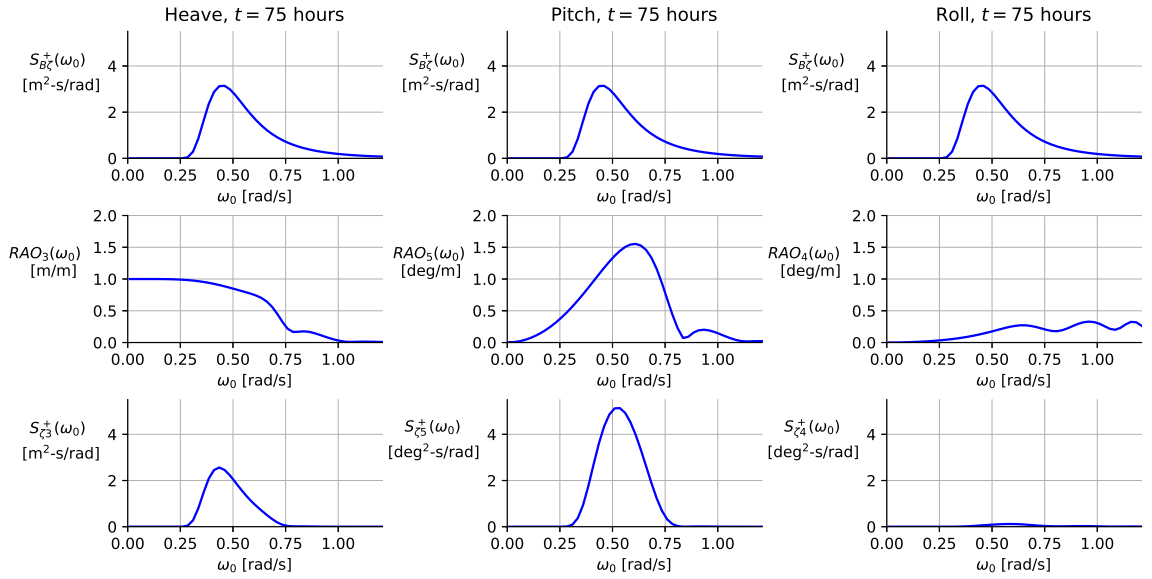

Figure 2.6: Calculation of the heave, pitch, and roll response spectra due to the wind wave partition at $t = 75$ hours using Equation 2.4 on a course from Guam to Hawaii. The top row indicates the wind wave input spectrum, the middle row shows the RAO for each DOF corresponding to the vessel's current speed and relative heading to the wind wave partition, and the bottom row shows the response spectra. As shown in Figure 2.8, the wind wave partition represents head seas at this time step, which provides significant heave and pitch excitation and minimal roll excitation.

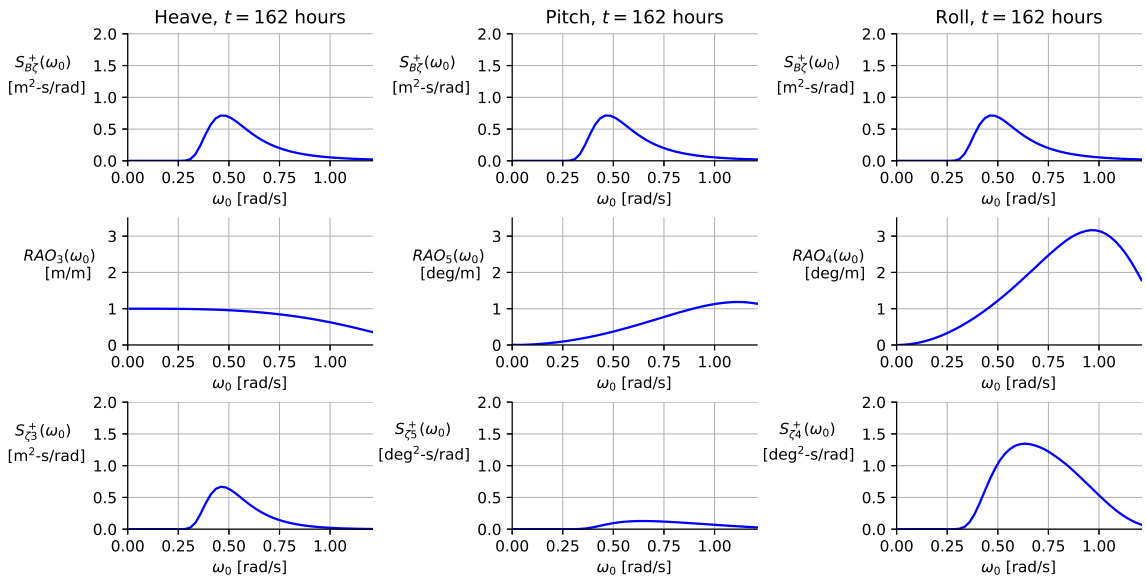

Figure 2.7: Calculation of the heave, pitch, and roll response spectra due to the wind wave partition at $t = 162$ hours using Equation 2.4 on a course from Guam to Hawaii. The top row indicates the wind wave input spectrum, the middle row shows the RAO for each DOF corresponding to the vessel's current speed and relative heading to the wind wave partition, and the bottom row shows the response spectra. As shown in Figure 2.8, the wind wave partition represents beam seas at this time step, which provides significant heave and roll excitation and minimal pitch excitation.

Equation 2.7, by showing the significant amplitudes in each DOF due to the wave partitions individually. It is important to note again that the linear superposition performed in the initial twin framework was of the wave partitions' zeroth spectral moments for a DOF and not a superposition of the significant amplitudes due to each wave partition, which is why the values of the significant amplitudes due to each wave partition in Figure 2.5 do not sum to equal the total significant amplitude. Figure 2.5 does, however, show the relative contributions of each wave partition to the total response. For example, the heave and pitch responses were dominated by the wind wave partition for a majority of the journey, while the roll response's greatest contributor was the primary swell for most of the voyage. These observed responses are greatly attributed to the directions of the incoming wave spectra, as explained in the following paragraphs.

Two demonstrations of the code's heave ($j = 3$), pitch ($j = 5$), and roll ($j = 4$)

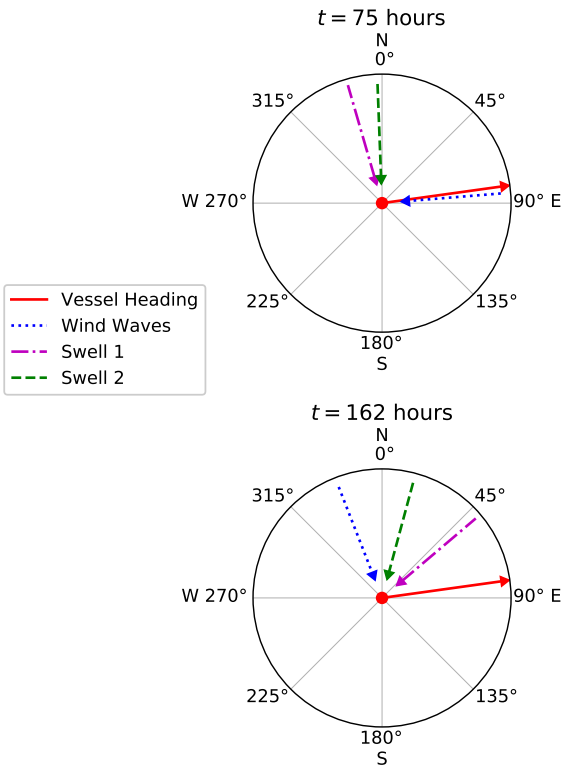

Figure 2.8: Compass plots for $t = 75$ hours and $t = 162$ hours corresponding to Figures 2.6 and 2.7, respectively.

response spectra calculations using Equation 2.4 are included in Figures 2.6 and 2.7, which reflect the vessel's response to the wind wave partition for forecast times of 75 hours, and 162 hours, respectively. The aforementioned values of j are commonly used DOF indices that are employed in this discussion. As shown in Figure 2.6, at $t = 75$ hours, the response spectra for heave and pitch are far more significant than the roll response spectrum. This difference is due to the fact that the wind wave partition for this time step was representative of head seas, which can be seen in Figure 2.8. Therefore, the wind wave partition did not significantly excite the roll DOF for this time step, but provided substantial excitation for heave and pitch. In contrast, in Figure 2.7, for $t = 162$ hours, the heave and pitch degrees of freedom have lesser relative responses to the wind wave partition than roll. This difference can also be explained by looking at the compass plots in Figure 2.8. For $t = 162$ hours, the wind wave partition represented beam seas, which excited roll greatly and provided

a notable contribution to heave, but yielded very little excitation in the pitch DOF.

The contrast between the two time steps shown in Figures 2.6 and 2.7 can be further explained by analyzing the heave, pitch, and roll significant amplitude plots in Figure 2.5. For $t = 75$ hours, the heave and pitch significant amplitudes due to wind waves were near their maximum values along the route because the vessel was encountering head seas with approximately the greatest wind wave significant amplitude seen during the voyage. The roll DOF, however, was not near its maximum value because roll is not excited by head seas as drastically as heave and pitch. For $t = 162$ hours, the roll significant amplitude was near its maximum for the voyage because the wind wave partition had shifted to beam seas. In contrast, the pitch significant amplitude due to wind waves was minimal because pitch is not excited by beam seas. Heave motion was apparent in both conditions, although, its amplitude was less significant for $t = 162$ hours. Although the vessel's crew may have perceived large-amplitude roll motions as being a more severe operation condition, the predicted fatigue damage accumulation rate at $t = 75$ hours was over 60 times greater than the fatigue damage accumulation rate at $t = 162$ hours. This drastic contrast emphasizes the correlation between large-amplitude pitch motions and fatigue damage accumulation, as well as the importance of fatigue damage monitoring.

2.3 Initial Twin Conclusions

The initial twin framework's fatigue damage estimation and tracking methods used wave data rather than relying on inspection results or expensive instrumentation aboard a vessel. The route from Guam to Hawaii showed significantly more damage per hour than the other three routes for the specific wave forecast data used. This increased accumulation rate was likely the result of the large head wind seas encountered along a majority of the route, which caused increased VBMs. If this trend continued and one ship accumulated significantly more fatigue damage over

months or years, responsibilities could be adjusted or traded with another vessel in the fleet to balance fatigue damage. This significant increase in cumulative fatigue damage for one of the four routes emphasized the potential benefits of fatigue damage tracking and balancing for support of life-cycle management decisions. This demonstration also showed the initial twin framework's potential for time-and-place specific response prediction, which would provide operators with forecasts of vessel motions rather than wave heights and directions alone to support underway and deployment decisions. Most importantly, this initial twin framework served as a foundation for improvement upon its traditional, physics-based approach in subsequent chapters using reflective twin concepts, which involved real-world measurements and data-driven models. Furthermore, this framework was necessary to demonstrate the importance of retaining physics-based information in data-driven models, which was achieved by including these PBMPs as input variables. As such, this framework's development was critical for all four research objectives outlined in Chapter I.

CHAPTER III

Wave Data Source Selection Study

While the initial twin framework demonstration of Chapter II revealed the potential benefits of wave-data-based response predictions and fatigue damage tracking for operational guidance and life-cycle management, it also prompted a critical question — What is the effect of wave data source selection on the output of the initial twin framework? If these wave data sources are to be used for owner and operator guidance in future, real-world implementations, understanding the impact of variation between wave data sources on the resultant response predictions and fatigue damage estimates is extremely important. The following study performed to address this question was published preliminarily in *Schirmann et al.* (2019b) with an expanded scope in *Schirmann et al.* (2020a).

Multiple agencies, such as NOAA's National Centers for Environmental Prediction (NCEP), which released the NWW3 wave forecast data used in the initial twin demonstration, release global wave forecast and hindcast data on a daily basis. These data sources are of interest to mariners and researchers alike and provide wave model output characterizing conditions at specific times and geographic locations, often defined by a uniform geospatial grid (by latitude and longitude) over a region. While comparisons of these wave models have been conducted in the past to benchmark performance for oceanographic modeling purposes among other reasons, published

studies on the consequences of these differences for vessel response prediction are limited. It is clear that significant differences between predicted and experienced wave conditions and resultant responses could be catastrophic for a vessel and its crew if sea state severity is underestimated; however, it is also important to consider that as the marine industry pushes toward digital monitoring and autonomous vessels, understanding these differences becomes increasingly important as human judgment is removed from the equation. Furthermore, as demonstrated in Chapter II, wave model data present an opportunity for virtual structural health monitoring using vessel position and these wave data alone, which may be a valuable, low-cost support mechanism for vessel owners, operators, and maintainers. Therefore, it was of utmost importance to conduct a formal investigation on the significance of which wave data source is selected for design-stage and through-life analyses of vessel motions, structural responses, and fatigue damage.

While the use of wave forecast or hindcast data to support operational, maintenance, and deployment decisions seems promising and beneficial, uncertainty in the wave data used may overshadow the benefits of this application. For example, while differences between wave hindcast datasets may seem insignificant when looking at minimal differences between resultant structural responses, fatigue damage estimation is a higher-order function of these responses, meaning the discrepancies between wave models may be amplified drastically. The primary goal of this study, discussed in *Schirmann et al.* (2019b) and *Schirmann et al.* (2020a), was to compare vessel motions, structural responses, and fatigue damage predicted using data from different hindcast wave data sources to determine the significance of wave data source selection for these analyses.

In addition to wave forecast and hindcast model data, although generally confined to coastal regions, a wide array of wave measurement buoys provide measured sea spectra that are also of interest to vessel operators and researchers, often as validation

of the wave models. Satellite measurements have also been employed for sea state characterization and in some cases, as data assimilated into wave hindcasts. Other work, such as that of *Nielsen* (2006), *Nielsen et al.* (2018), and *Duz et al.* (2019), has focused on sea state estimation using vessel motions (i.e., ship as a wave buoy approaches). *Nielsen et al.* (2019) discussed a framework for extension of ship as a wave buoy approaches to incorporate data from multiple vessels.

There have been numerous comparison studies on the differences between data from forecast and hindcast wave models, as well as real-world measurements from buoys and satellites; however, these comparison studies focus on wind and wave conditions rather than vessel motions and structural responses. For instance, *Stopa and Cheung* (2014) compared wind speeds and wave heights between different weather models using buoy data and satellite altimeters for several geographically diverse regions, and *Campos and Guedes Soares* (2016) compared three wave hindcasts with satellite data over a large region in the North Atlantic Ocean. *Guedes Soares and Moan* (1991) also examined the effect of uncertainty in wave statistics on fatigue damage, showing that wave data source selection has been an ongoing concern in the marine industry for decades. Rather than further validation of existing wave models or investigation of the differences between them from an oceanographic standpoint, the goal of this study was to examine the effects of these differences on the initial twin framework's vessel response predictions and fatigue damage estimates.

In this study, vessel responses and fatigue damage estimates resulting from different hindcast wave data sources were compared at two locations off the coast of Alaska and one location near Hawaii. The wave data compared in this work came from NOAA's WAVEWATCH III (NWW3) Multigrid Production Hindcast (*NOAA National Centers for Environmental Prediction.*, 2019), the EU's Copernicus Marine Environment Monitoring Service (CMEMS) Global Ocean Waves Analysis and Forecasting Product (*Fernandez and Aouf*, 2018), and the National Data Buoy Center

(*NOAA's National Data Buoy Center Center of Excellence in Marine Technology*, 2018). Using these different wave data sources, the initial twin framework's heave, pitch, VBM, and fatigue damage predictions for a ship operating with an array of speed and heading combinations were compared for the full year of 2017. Although the most compelling findings of *Schirmann et al.* (2019b) and *Schirmann et al.* (2020a) presented in this chapter related to VBMs and fatigue damage for virtual structural health monitoring, heave and pitch were included to show the implications of wave data differences for operational guidance. Furthermore, in comparison to heave and pitch, VBMs are difficult and expensive to measure onboard a vessel. As such, establishing a similarity between the influence of wave data differences on motions and structural responses may prove beneficial for future studies. The comparison approach employed in this chapter highlights the impact of wave data selection on ship response predictions and fatigue estimates without addressing uncertainties from ship modeling itself (e.g., 2D strip-theory idealization, locations of the vertical and longitudinal centers of gravity, etc.).

3.1 Methodology

This section discusses the approach employed in *Schirmann et al.* (2019b) and *Schirmann et al.* (2020a) to investigate the significance of wave data source selection for vessel response prediction and fatigue damage estimation. Section 3.1.1 discusses the three wave data comparison locations, and Section 3.1.2 discusses the simulations performed at each location. Sections 3.1.3 and 3.1.4 discuss the wave model and buoy data considered at each location, respectively, and the breadth of this study is summarized in Section 3.1.5. Section 3.1.6 describes the the initial twin framework's application for response prediction and fatigue damage estimation in this study, and Sections 3.1.7 and 3.1.8 discuss the metrics and parameters used to compare wave data sources.

3.1.1 Comparison Locations

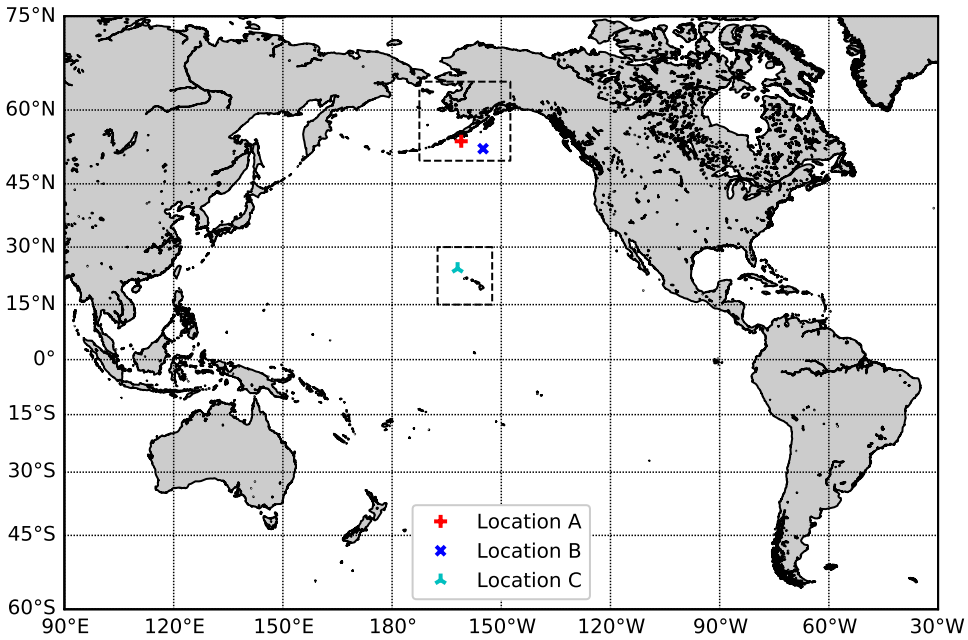

Figure 3.1: A broad overview of the three locations within the Pacific Ocean that were considered in this work.

Three locations were selected for comparison of wave data from different sources, as well as vessel response predictions and cumulative fatigue damage estimates generated using these wave data. Two locations were selected off the coast of the Aleutian islands in Alaska, and the third location was chosen near the northwestern Hawaiian islands. Figure 3.1 gives a broad overview of both regions and the three locations within them in the Pacific Ocean. The two regions identified in Figure 3.1 are enhanced in Figure 3.2. These two regions were chosen in part because they had wave data available from multiple wave data sources at varying geospatial grid resolutions and had potential to experience severe sea states that would add to the breadth of this study. Within the two regions, these three specific locations were selected because they were close in proximity to buoys managed by the National Data Buoy Center (NDBC) with measured directional wave spectra available. Additionally, the geospatial output grids, defined as the locations where wave hindcast data were avail-

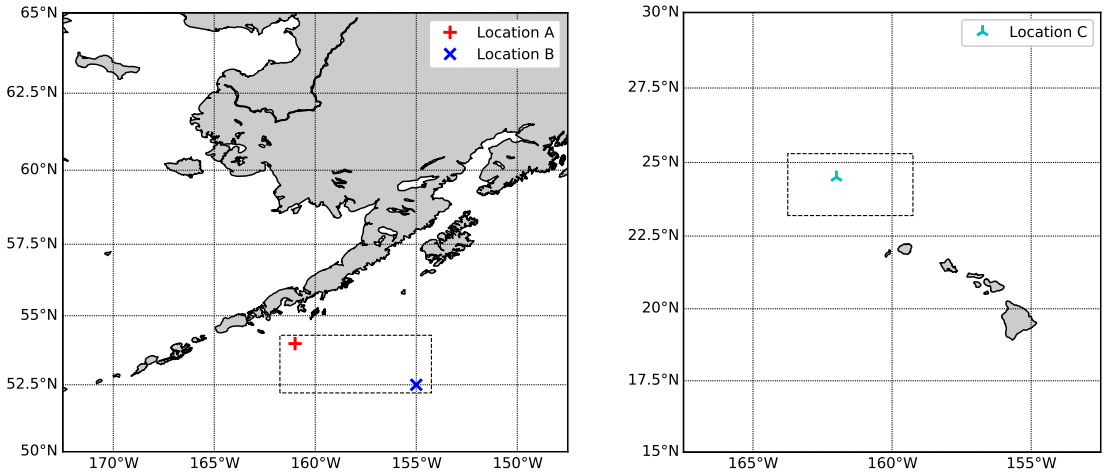

Figure 3.2: Outlined regions from Figure 3.1 showing the two Alaskan and one Hawaiian wave data comparison locations. The outlined regions are enhanced below in Figure 3.3 to show the locations of the NDBC buoys relative to wave model output Locations A, B, and C.

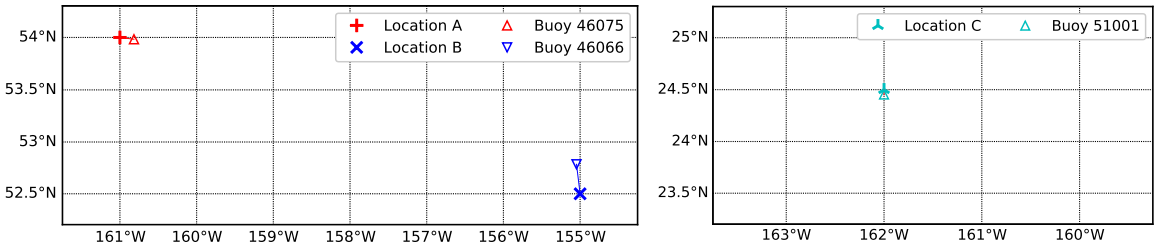

Figure 3.3: Enhanced map regions outlined in Figure 3.2 to show the locations of the NDBC buoys relative to wave model output Locations A, B, and C.

able, from multiple wave data sources overlapped exactly at these locations. The three analysis locations (where the considered wave model grids overlapped) and the locations of the nearby NDBC buoys are described in Table 3.1. Table 3.1 includes the approximate distance between each analysis location and the nearby buoy's location, which was calculated using `https://www.gpsvisualizer.com/calculators`. The outlined regions of Figure 3.2 are enhanced further in Figure 3.3 to show the placement of the buoys relative to the three analysis locations.

Table 3.1: Wave data comparison locations, nearby buoys, and the approximate distance between the buoys and wave data locations (from `https://www.gpsvisualizer.com/calculators`).

	Location A	Location B	Location C
NDBC Buoy ID	46075	46066	51001
Wave Model Location	54.00°N, 161.00°W	52.50°N, 155.00°W	24.50°N, 162.00°W
NDBC Buoy Location	53.98°N, 160.82°W	52.79°N, 155.05°W	24.45°N, 162.00°W
Distance to Buoy	7 NM	18 NM	3 NM

3.1.2 Simulations Performed

The destroyer-sized naval combatant used in the initial twin demonstration of Chapter II (Table 2.1) and described in *Stern et al.* (2011), the *DTMB 5415*, was placed at each location for the full year of 2017 (i.e., from January 1 - December 31, 2017). Vessel response simulations were carried out using all combinations of four speeds and eight compass headings, which resulted in 32 total cases for each given time, location, and wave data source. The four speeds tested were 0, 10, 20, and 30 knots, and the eight headings were evenly spaced on $[0^{\circ}, 315^{\circ}]$ in increments of 45°. In each of these 32 test cases, heave, pitch, and VBM responses were predicted using the initial twin framework for every three-hour time increment during the one-year exposure time. Three-hour intervals were selected to ensure data availability from all wave data sources at each time step. Due to the large volume of data and test cases considered, this work was performed using a high-performance computing cluster provided by Advanced Research Computing at the University of Michigan, Ann Arbor. As mentioned in Chapter II, the predicted VBM responses were used to estimate the fatigue damage accumulated amidships over time. The wave data and resultant predicted responses and fatigue damage estimates from different wave data sources were then compared.

3.1.3 Wave Model Data

Four NWW3 hindcast datasets and one CMEMS hindcast dataset were used in this study. The availability of each dataset at the three locations is indicated in

Table 3.2: Wave data source availability (✓) at each location. The ✗ symbol indicates that the dataset was available but not considered due to findings at the other locations. Note that the figures of Section 3.2 only include a single NWW3 source for each location unless otherwise stated.

Wave Data Source	Location A	Location B	Location C
NWW3 Glo 30 min		✓	✗
NWW3 AK 10 min	✓	✓	
NWW3 EP 10 min			✓
NWW3 AK 4 min	✓		
CMEMS Glo 5 min	✓	✓	✓
NDBC Buoy (Measured Spectra)	✓	✓	✓
NDBC Buoy (Bretschneider)		✓	

Table 3.2. These datasets had varying geospatial resolutions which are indicated in the dataset name (i.e., NWW3 Glo 30 min had an approximate latitude/longitude resolution of 30 arcminutes or 0.5°). Note, however, that these resolutions were approximate and not always consistent between latitude and longitude; due to the actual geospatial increments and different masking patterns in the datasets, it was only possible to overlap two of the considered NWW3 datasets and the one CMEMS dataset at Locations A and B. Therefore, one NWW3 dataset was exchanged for another when shifting the analysis from Location A to Location B. Table 3.2 also notes that while NWW3 Glo 30 min data were available at Location C, they were not considered in this study due to the preliminary findings for Locations A and B presented in *Schirmann et al.* (2019b), which showed that NWW3 grid resolution did not matter. These results are revisited in the following discussion.

3.1.3.1 NWW3 Multigrid Production Hindcast

As shown in Table 3.2, the four NWW3 grids used, in order of increasing spatial resolution, were the Global 30 arcminute grid (NWW3 Glo 30 min), Alaskan 10 arcminute grid (NWW3 AK 10 min), the East Pacific 10 arcminute grid (NWW3 EP 10 min), and the Alaskan 4 arcminute grid (NWW3 AK 4 min). Note that within the same region (e.g., near Alaska), as spatial resolution increased, the geographical area

with available data decreased. The NWW3 Glo 30 min grid spatial domain included a majority of the world's oceans, while the NWW3 AK 10 min grid was masked to cover a large coastal region off the coast of Alaska, and the NWW3 AK 4 min grid covered a less expansive area off the coast of Alaska. The NWW3 EP 10 min grid covered a large region off the western coast of the United States, as well as a separate region surrounding the Hawaiian island chain that contained Location C. The NWW3 Multigrid Production Hindcast that provided these datasets used the NCEP's Global Forecast System (GFS) winds as input forcing to the multi-grid spectral wave model WAVEWATCH III (*NOAA National Centers for Environmental Prediction.*, 2019). There was no wave data assimilation from wave buoys, satellites, or other sources included in these data, meaning no real-world wave data were used to correct the hindcast output. At a given time and location, the NWW3 datasets characterized the wave conditions using a varying number of independent, directional wave partitions, each defined by a significant wave height, peak period, and mean direction. Note that the number of wave partitions defined for a given time and location was often greater than three, which is a notable difference from the NWW3 wave data used in the initial twin framework demonstration of Chapter II. Detailed information about the NWW3 partitioning scheme can be found in *The WAVEWATCH III® Development Group (WW3DG)* (2016).

3.1.3.2 CMEMS Global Ocean Waves Analysis and Forecasting Product

The single CMEMS grid analyzed in this study, referred to as CMEMS Glo 5 min for consistency with the NWW3 abbreviations, came from the Global Ocean Waves Analysis and Forecasting Product (GLOBAL_ANALYSIS_FORECAST_WAV_001_027). This dataset covered a majority of the world's oceans with an approximate geospatial resolution of 5 arcminutes and used winds and ice fields from the European Centre for Medium-Range Weather Forecasts (ECMWF) to force the Meteo-France

Wave Model (MFWAM) (*Fernandez and Aouf*, 2018). Significant wave height data were assimilated in this wave model from satellite altimeters. Similar to the NWW3 data employed in Chapter II, field output wave data for a given time and location were provided as three independent, directional wave partitions: wind waves, primary swell, and secondary swell. To maintain consistency between different datasets, the specific classification of each wave partition type was not used to differentiate between partitions in this analysis (i.e., the assumed spectral shape used in the initial twin framework was not modified based on partition type). Each of the three wave partitions from CMEMS Glo 5 min was characterized by a significant wave height, mean period, and mean direction.

3.1.3.3 NWW3 and CMEMS Idealized Wave Spectra

For all of the wave model datasets used in this work, a Bretschneider (ITTC two-parameter) spectrum was constructed for each independent, directional wave partition at a given time step and location using the provided significant wave height and period. As in Chapter II, this step was necessary because only characteristic parameters were provided for each wave partition rather than fully defined spectra. Note that the Copernicus Climate Change Service does provide hindcast 2D wave spectra that are used in Chapters VI and VIII of this dissertation. However, to limit the scope of this wave data study, only one data source with 2D spectra was used, which was the measured wave buoy data described in the following section. Nevertheless, comparisons between the Copernicus data options (2D spectra and CMEMS directional wave partitions) are worthy of further study.

3.1.4 Wave Buoy Data

The measured NDBC buoy wave data for a given time step and location were provided as fully defined 1D spectra with coefficients included for a directional Fourier

series expansion to appropriately spread the spectral energy on $[0, 2\pi)$ radians. The directional spreading function using the four parameters (R_1, R_2, α_1, and α_2), which were provided in the datasets for each discrete frequency, f, is defined in Equation 3.1.

$$D(f, \alpha) = \frac{1}{\pi}\left(\frac{1}{2} + \frac{2}{3}R_1 \cos(\alpha - \alpha_1) + \frac{1}{6}R_2 \cos\left[2(\alpha - \alpha_2)\right]\right) \tag{3.1}$$

In Equation 3.1, α is the direction that the waves come from measured clockwise from true north, and integration of $D(f, \alpha)$ on $[0, 2\pi)$ radians yields a value of 1.0. Note that the weighting terms ($\frac{2}{3}$ and $\frac{1}{6}$) discussed in *Earle et al.* (1999) are incorporated in Equation 3.1, which *NOAA's National Data Buoy Center Center of Excellence in Marine Technology* (2018) presented as an option to avoid negative values in the spreading function through smoothing. As noted in *Earle et al.* (1999), without these weighting terms, the directional resolution (half-power width) of this spreading function would be 88°. With these weighting terms, this directional resolution metric is increased to 130°.

After transforming the 1D wave spectrum using the provided spreading parameters for each frequency, the resulting 2D spectrum was partitioned into eight evenly spaced bins covering 45° each. Each of the eight bins was defined by a 1D wave spectrum and the mean direction within the bin, both resulting from integration over the bin's directional spread. For the purposes of this analysis, the eight bins were then treated as eight directional wave partitions for calculation of the vessel's responses. In contrast to the wave model data, because the measured spectral shape associated with each directional wave partition was known for the buoy data, construction of a Bretschneider spectrum was not necessary; however, to investigate the importance of spectral shape on these analyses, an additional experiment was completed at Location B in which the mean period and significant wave height of each of the eight bins were used to generate a Bretschneider spectrum. While peak period could have been used to construct the idealized spectrum, the measured spectrum in a single bin was not

necessarily unimodal, so mean period was selected. Nevertheless, either period metric would have yielded a change in spectral shape relative to the measured data, which was the desired effect. This additional experiment at Location B is indicated in Table 3.2. As in the other experiments, simulations in all 32 heading/speed combinations were completed for the full one-year period, and the resultant response predictions and fatigue damage estimates were compared to the buoy data results using measured spectra, as well as the wave model data results.

3.1.5 Summary of Study Breadth

As outlined in Sections 3.1.2–3.1.4 and summarized in Table 3.2, a minimum of three wave data sources and their resultant responses and fatigue damage were compared at each of the three locations for a given three-hour time increment of the one-year period. As shown by the check marks (✓) in Table 3.2, between the three locations, 12 different wave data sources, including repeats between locations and variants of the sources (i.e., Bretschneider reconstruction of the buoy data), were tested. For each of these 12 wave data source instances considered and 2,920 three-hour time increments in the year 2017, 32 speed and heading combinations were simulated to generate 3 types of response predictions: heave, pitch, and VBM. Combination of these numbers ($12 \times 2,920 \times 32 \times 3$) yields a total of 3,363,840 responses predicted in this study. As mentioned in Section 3.1.2, these response prediction calculations were performed using the initial twin framework and a high-performance computing cluster.

3.1.6 Response Prediction and Fatigue Damage Estimation

As mentioned previously, the wave data from all sources were analyzed in three-hour time increments. A required assumption to obtain cumulative fatigue damage from the initial twin framework's frequency-domain approach was that the wave en-

vironment and operating conditions were stationary over the entire three-hour period (from 1.5 hours before to 1.5 hours after the specified wave data time). During real-world operation of a ship, forward speed would lead to variation in the vessel's location during this period. As a result, even if conditions were temporally stationary, the wave conditions seen by the vessel could change. The primary goal of this work was not to simulate a ship at sea for a whole year as a realistic demonstration, but rather to make controlled comparisons between wave data sources and their resultant responses and fatigue damage. While forward speed was considered for the sake of response prediction and fatigue damage estimation, the vessel's physical location was assumed constant over the three-hour window. This constraint is analogous to a scenario in which a ship travels within a region that is temporally and spatially stationary during the three-hour period.

As in Chapter II, the linear, 2D strip-theory code developed by *Beck and Troesch* (1990), `SHIPMO.BM`, was used to calculate RAOs for this study. Note that nonlinear response calculations would be expected to yield higher bending moments in some of the more-severe sea states encountered, with corresponding increases in fatigue damage. Heave, pitch, and VBM RAOs were calculated at all four speeds and in 5° relative heading increments on $[0^\circ, 355^\circ]$. For each wave partition, the RAO with the closest relative heading to the calculated relative heading (between the wave partition and ship) was used. Directional interpolation between RAOs was deemed unnecessary because the 5° relative heading increments yielded minimal differences between RAOs. Note that wave spreading functions were not applied to any of the wave partitions from any of the wave data sources; this decision was made to avoid unfairly influencing any of the wave data sources with arbitrary spreading functions prior to comparisons.

The vessel response comparisons in the following discussion were made using the response significant amplitudes, $\zeta_{j,T}$, averaged over all eight headings for a given loca-

tion, time, wave data source, and speed. While directional averaging was necessary to make the following analysis and discussion tractable given the large amount of data, the implications of this decision are discussed in Sections 3.2.2 and 3.2.3.

The purpose of the fatigue analysis performed in this work was not to exactly model a specific structural detail, but instead to have a representative calculation for comparison between wave data sources. As such, fatigue damage was estimated at Station 10 (amidships) where the highest VBM responses were witnessed in the demonstration of Chapter II. The same bottom cross section modulus of 3.77 m^3 from *Ashe et al.* (2009) was used in this study. In the preliminary work of this study presented in *Schirmann et al.* (2019b), the narrow-banded fatigue damage approximation described in Section 2.1.3.1 was used to simplify the analysis. In the expanded study published in *Schirmann et al.* (2020a), the improved fatigue damage estimation method of Section 2.1.3.2 was employed. As a result of this improved fatigue damage estimation procedure and adjustments in the wave data used for response prediction, some of the preliminary results presented in *Schirmann et al.* (2019b), which only included Locations A and B, changed.

The vessel's cumulative fatigue damage values resulting from the different wave data sources over the one-year period were compared. These comparisons were made separately for each location and speed using the average damage value over all eight ship headings. As mentioned previously, the implications of direction averaging are revisited in Sections 3.2.2 and 3.2.3.

3.1.7 Comparison Methods

Although several comparison metrics were considered and tested, one metric, monthly mean bias, was selected for presentation to quantify the differences between wave data sources and their resultant heave, pitch, and VBM responses. Bias was selected in particular because it preserves the sign of the differences between wave

data sources. Monthly mean bias, $\overline{BIAS}$, was calculated using Equation 3.2, which was adapted from a bias equation used in *Campos and Guedes Soares* (2016) for comparison of wave hindcast data and satellite measurements. Analyzing the data on a monthly basis allowed potential seasonal variation in the relationships between data sources to be uncovered.

$$\overline{BIAS} = \frac{1}{N}\sum_{i=1}^{N}(M_i - B_i) \tag{3.2}$$

In Equation 3.2, M_i was the value from or resulting from a given wave model data source (NWW3 or CMEMS) at time step i, and B_i was the value from or resulting from the nearby buoy data (using measured spectra) at time step i. When comparing the Bretschneider reconstruction of the buoy data to the the measured buoy data at Location B, the Bretschneider reconstruction results were used as M_i. For heave, pitch, and VBM responses, separate monthly mean bias values were calculated for each speed. At a given speed, the differences between M_i and B_i were averaged over all N three-hour time steps for all eight directions considered in the month (i.e., $N = 8 \times$ number of three-hour time intervals in the month). When comparing mean wave directions, the absolute value was taken in the numerator of Equation 3.2 to yield the monthly mean absolute bias because the sign of the difference was not significant for direction. Note that although the buoy was treated as the foundation for the comparisons made using Equation 3.2, there was no intention to suggest that the buoy measurements were the ground truth or even that they were more accurate than the wave model data; the goal of this study was not to validate or find fault with any of the tested wave data sources, but rather to analyze the differences between these sources and the effects on the initial twin framework's response predictions and fatigue damage estimates.

3.1.8 Wave Data Comparison Parameters

Five characteristic parameters were used for wave data comparison in the following analysis: significant wave height, mean period, spectral peak period, mean direction, and bandwidth. For a given time, location, and wave data source, the total wave spectrum was calculated by summing the independent wave partition spectra at each wave frequency. The significant wave height and mean period were then calculated using conventional methods from this total wave spectrum. The spectral peak period was the wave period corresponding to the maximum value in the total wave spectrum, which is an indicator of the dominant wave partition period. The mean direction was determined by taking the average of individual wave partition mean directions weighted by the zeroth moment of the wave partition spectra. These mean direction calculations were performed using vectors to avoid issues with circularity. The wave direction data presented in the following discussion are defined as the direction of wave propagation with 0° and 90° corresponding to North and East, respectively. The bandwidth of the wave spectrum was calculated following Equations 2.5, 2.6, and 2.14. These five characteristic values allowed direct comparisons of wave data sources with differing data formats.

3.2 Results and Discussion

The following section compares wave data, resultant responses, and fatigue damage estimates from different wave data sources at three different locations as presented in *Schirmann et al.* (2020a). As mentioned previously, some of the results and discussion for Locations A and B were adapted from the preliminary study of *Schirmann et al.* (2019b); however, some of these previously discussed results were updated due to the improved fatigue damage estimation procedure and adjustments in the wave data used for response prediction in the initial twin framework.

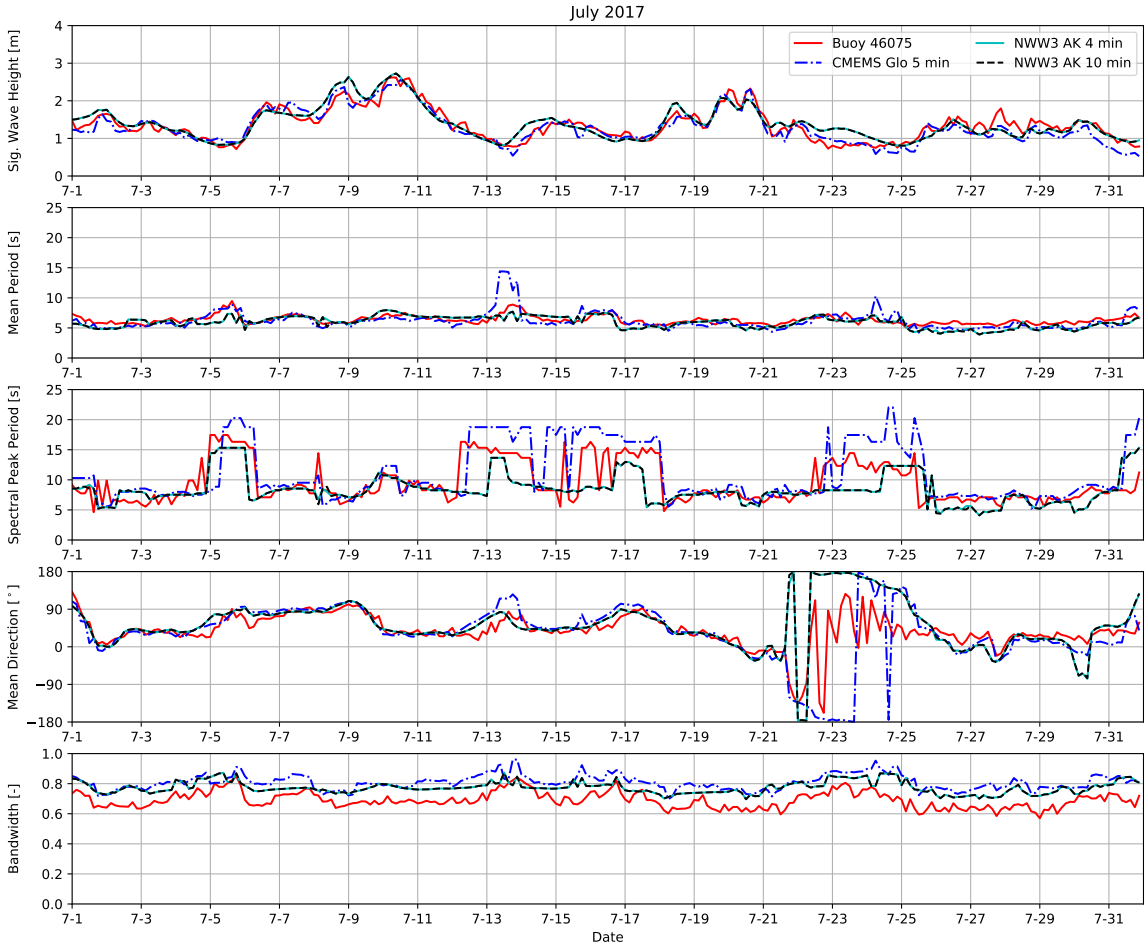

Figure 3.4: July 2017 time series of significant wave height, mean period, spectral peak period, mean direction, and bandwidth at NDBC Buoy 46075 and Location A for NWW3 AK 4 min, NWW3 AK 10 min, and CMEMS Glo 5 min. The NWW3 AK 4 min and AK 10 min wave data curves overlap nearly perfectly. Mean direction values indicate the direction of wave propagation and were plotted on [-180, 180)° rather than [0, 360)° to limit fluctuations at the y-axis boundaries. Following standard compass directions, 0° and 90° correspond to North and East, respectively.

3.2.1 Wave Data Comparisons

A time series for July 2017 showing the significant wave height, mean period, spectral peak period, mean direction, and bandwidth of the input wave data at Location A is presented in Figure 3.4. July 2017 was selected because it had the lowest monthly mean significant wave height measured at NDBC Buoy 46075 for the year, meaning wave conditions were relatively benign. The NWW3 AK 10 min and NWW3 AK 4

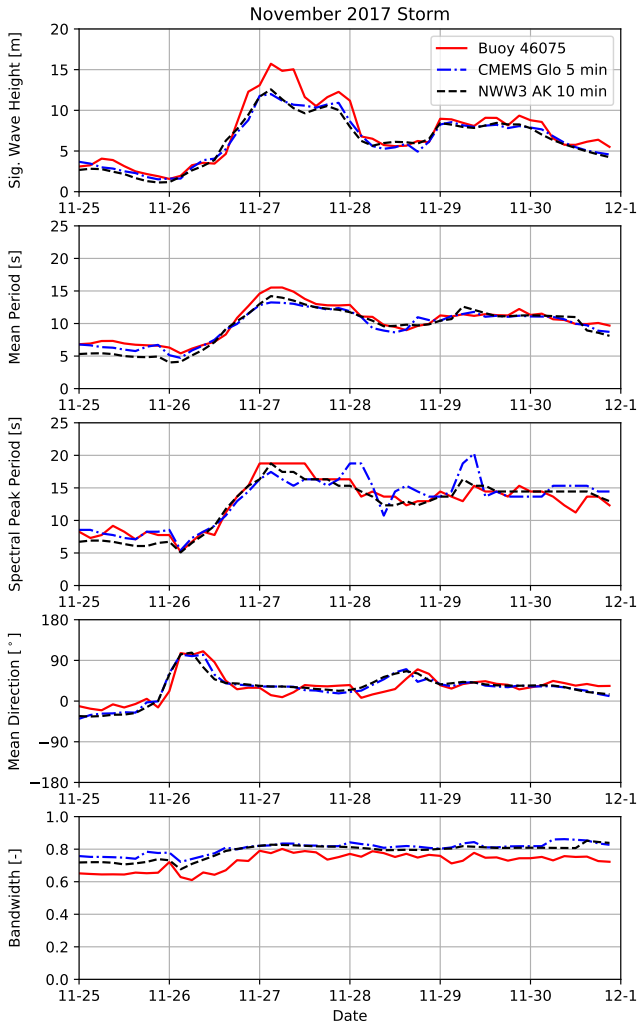

Figure 3.5: November 2017 storm time series of significant wave height, mean period, spectral peak period, mean direction, and bandwidth for NWW3 AK 10 min, CMEMS Glo 5 min, and Buoy 46075. Mean direction values indicate the direction of wave propagation and were plotted on [-180, 180)° rather than [0, 360)° to limit fluctuations at the y-axis boundaries. Following standard compass directions, 0° and 90° correspond to North and East, respectively.

min plots overlap nearly perfectly, which implies that spatial resolution was not a large factor in the NWW3 data sets for Location A specifically. This nearly identical agreement continued for the entire year, which also yielded nearly identical response predictions and fatigue damage estimates. It was found that this agreement also held true between NWW3 AK 10 min and NWW3 Glo 30 min at Location B. It is possible that locations closer to the coast or in different geographical regions would

yield larger differences between NWW3 spatial resolutions, and the higher resolution grids include points that are not included in the other data sets which are useful for near-shore analyses. Nevertheless, at these two locations, the differences in spatial resolutions of the NWW3 grids did not have a significant effect; therefore, in the following discussion, only the AK 10 min grid is presented for Locations A and B. The AK 10 min wave data were selected as the representative NWW3 data set because they were available at both analysis locations. As indicated in Table 3.2, due to the findings at Locations A and B, only one NWW3 grid, EP 10 min, was analyzed at Location C.

As shown in the July 2017 time series of Figure 3.4, while the significant wave heights from NWW3, CMEMS, and the buoy appear to show reasonably strong agreement, there were times in which the curves deviated from each other even in these benign conditions. The mean period and direction data from all three sources also agreed reasonably well, excluding a few time steps where the buoy appears to be approximately 180° out of phase with the two models, but the spectral peak periods of these data sets included stretches of notable disagreement. As mentioned in Section 3.1.8, the peak period is an indicator of the dominant wave partition period. Some areas of disagreement in the spectral peak periods between wave data sources may have occurred due to time lags in wave system arrivals to the analyzed location. Additionally, the differences in spectral peak periods may have resulted from the required assumption of spectral shape (Bretschneider) for the wave model data, which was also influenced by the wave period definition (mean versus peak) provided by the wave data source. The bandwidth time series shows that the measured buoy data typically had a lower bandwidth than the wave models, which was unexpected because the wave models were constructed using narrow-banded Bretschneider spectra for each directional wave partition. Although complex and not easily quantifiable, the effect of these peak period and bandwidth discrepancies on the predicted responses and

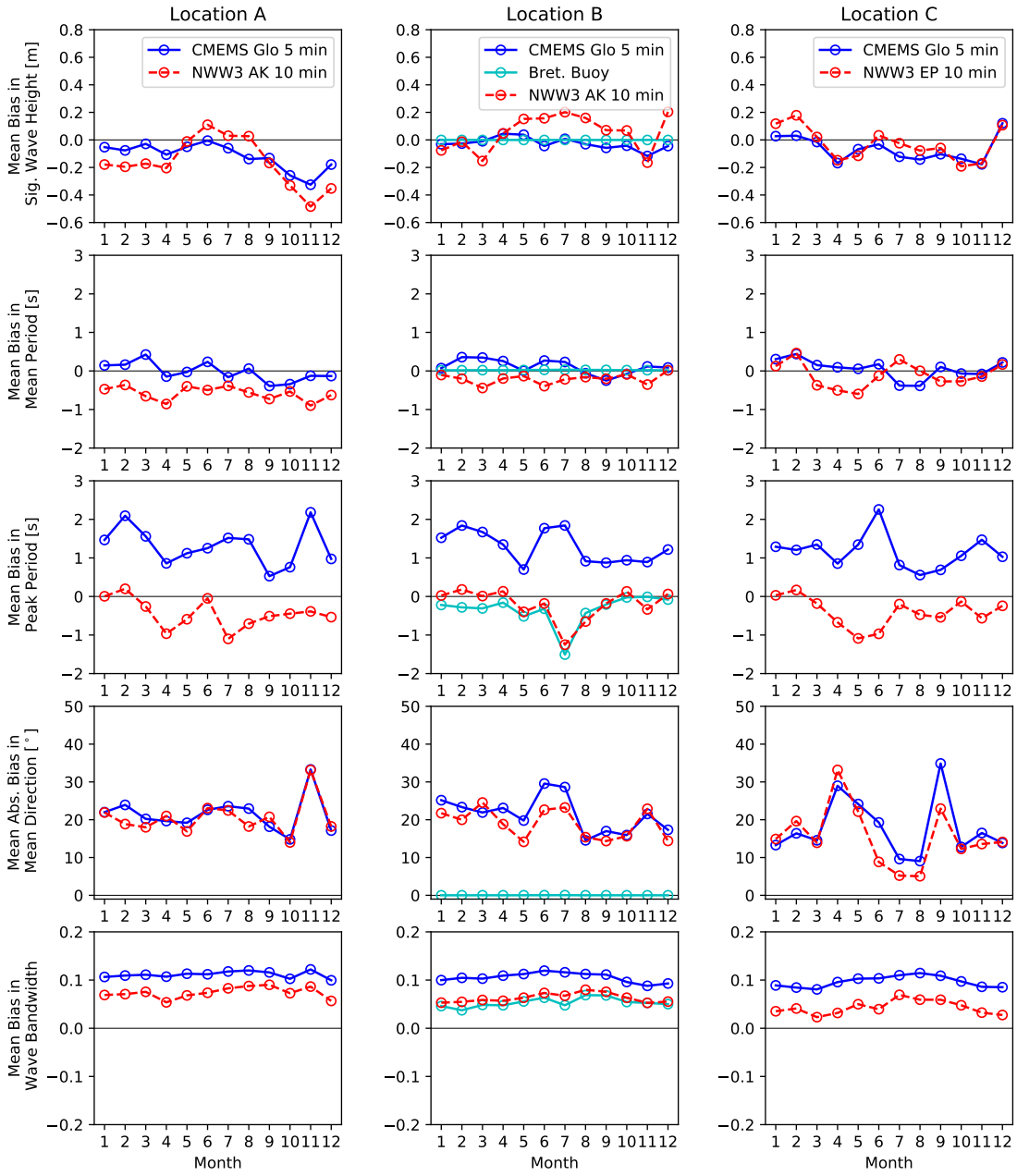

Figure 3.6: Monthly mean bias in significant wave height, mean period, spectral peak period, mean direction, and bandwidth at Locations A, B, and C. Location B also includes the Bretschneider reconstruction of the buoy data.

fatigue damage estimates may have been significant.

A time series of the five wave data parameters for a November 2017 storm with severe wave conditions at Location A is shown in Figure 3.5. Once again, NWW3 and CMEMS show reasonably strong agreement; however, at the highest significant

wave heights seen, NWW3 and CMEMS underestimated the significant wave height relative to the buoy data. Better agreement was present in the spectral peak period here than previously in the July time series. This agreement is largely attributed to the selected period metric because these large sea states had a high likelihood of being dominated by a single wave partition. Once again, the measured buoy data appear to be more narrow-banded than the NWW3 and CMEMS wave model data. The mean wave periods and directions show reasonably strong agreement for the entire storm.

Plots of the monthly mean bias in significant wave height, mean period, spectral peak period, mean direction, and bandwidth calculated using Equation 3.2 are shown in Figure 3.6 for Locations A, B, and C. Location B also includes the monthly mean bias plots of the Bretschneider reconstruction, which was an idealized spectrum created using the measured significant wave height and mean period from the buoy. As a result, all the bias values of significant wave height, mean period, and mean direction for the Bretschneider reconstruction were zero; however, the peak period and bandwidth bias plots show notable disagreements between the Bretschneider reconstruction and the buoy data because these parameters were not preserved in the reconstruction. Note that if the reconstruction was performed using peak period instead of mean period, the peak period bias values of the Bretschneider reconstruction would be zero and the mean period bias values would be nonzero.

In Figure 3.6, the monthly mean bias values in significant wave height at all three locations show that subtle disagreement existed between the wave data sources used to generate response predictions and fatigue damage estimates within the initial twin framework. Figure 3.6 also reveals that while the mean periods of all three wave data sources agreed reasonably well, CMEMS tended to overestimate the spectral peak period relative to NWW3 and the buoy, which may also have resulted from the period definition differences between NWW3 and CMEMS datasets. At Location B, the peak period bias values of the Bretschneider buoy reconstruction and the

NWW3 datasets were very similar, which is interesting because peak period was used to construct the NWW3 idealized spectra while mean period was used for the buoy reconstruction. The mean absolute bias in mean direction shows strong agreement between NWW3 and CMEMS in general, but both absolute bias values indicate a difference relative to the buoy. This difference reinforces the fact that the buoy data should not have necessarily been treated as the ground truth. As mentioned in Section 3.1.4, the directional resolution (half-power width) of the buoy data used in this study was 130°, which may have contributed to the directional discrepancies between the buoy and wave model data. As shown by the differences outlined here, wave data source selection for response prediction and fatigue damage estimation should not be trivialized.

3.2.2 Response Comparisons

Time series plots of the direction-averaged heave, pitch, and VBM significant amplitudes for July 2017 are shown in Figure 3.7 for a speed of 20 knots at Location A. The significant wave height time series that was originally shown in Figure 3.4 is also included here for comparison purposes. It can be seen in Figure 3.7 that the differences between significant wave heights from each wave data source propagated to the vessel responses. Note that while only the 20 knots responses are shown, this propagation was also evident for the other three forward speeds, as well as at Locations B and C. The direction-averaged response time series plots for a speed of 10 knots in the November storm examined previously are shown in Figure 3.8, which also shows that relationships between significant wave heights from different data sources propagated to all three responses. While significant wave height is included in Figures 3.7 and 3.8 for reference, the discrepancies between responses were also the result of differences in the other four wave parameters of Figures 3.4 and 3.5.

Monthly mean bias plots for direction-averaged heave, pitch, and VBM signifi-

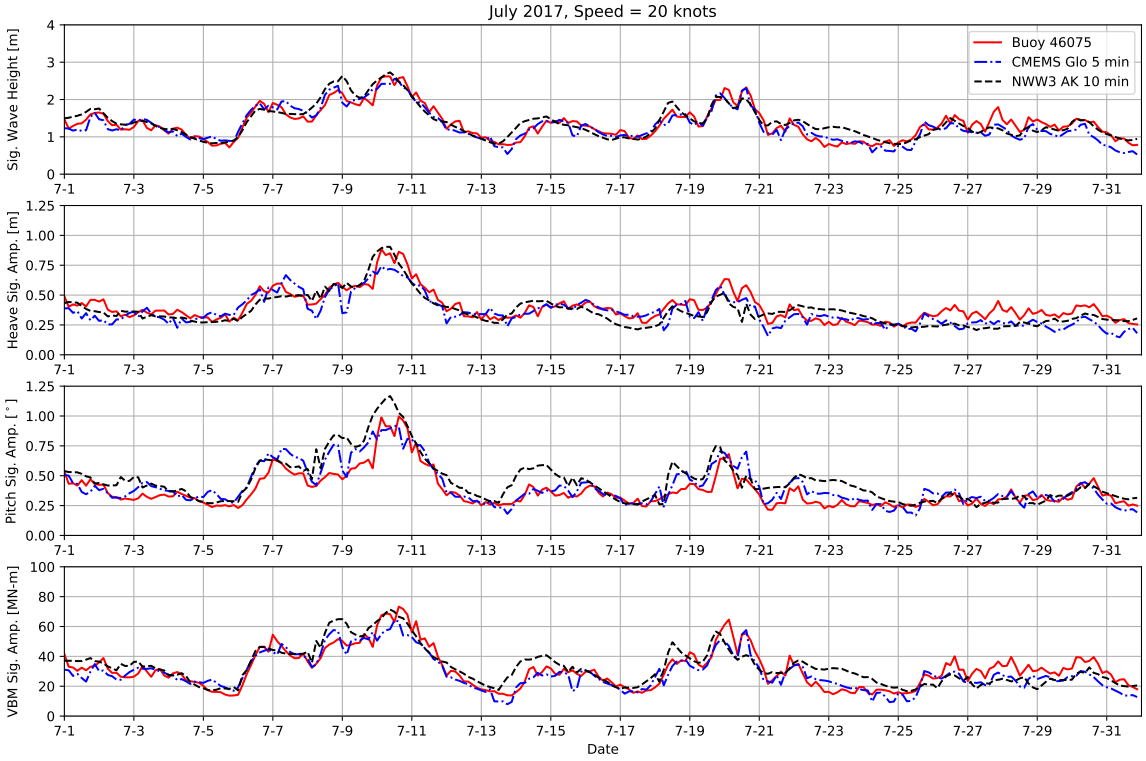

Figure 3.7: July 2017 time series of significant wave height and the significant amplitudes of heave, pitch, and VBM responses for a speed of 20 knots at NDBC 46075 and Location A for NWW3 AK 10 min and CMEMS Glo 5 min.

cant amplitudes are shown in Figure 3.9. The four speeds are plotted as different markers, and the lines indicate the monthly mean bias over all speeds. As in Figure 3.6, Location B includes the Bretschneider reconstruction of the buoy data. The significant wave height mean bias values from Figure 3.6 are also included for comparison. In Figure 3.9, the significant wave height biases of NWW3 and CMEMS propagated to the response significant amplitude biases. For all wave data sources and locations, variation in speed did not have a significant impact on bias, excluding VBM predictions at 30 knots.

As mentioned in Section 3.1.6, while direction-averaging was necessary to make this discussion tractable, there was certainly variation in response magnitudes and the resultant bias values as the vessel's relative heading to the wave partition(s) changed. For example, in general, predominant head seas yield a much higher pitch

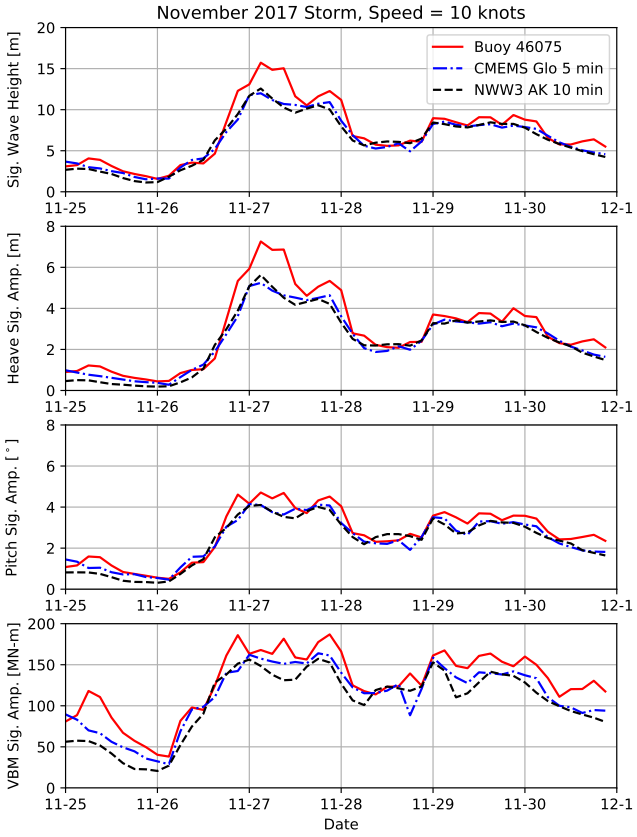

Figure 3.8: November 2017 storm time series of significant wave height and the significant amplitudes of heave, pitch, and VBM responses for a speed of 10 knots at Buoy 46075 and at Location A for NWW3 AK 10 min and CMEMS Glo 5 min.

and VBM response than predominant beam seas, as demonstrated in Chapter II. As a result, large bias values in predominant head seas may not be adequately portrayed by the direction-averaged values if the beam-seas responses from all wave data sources were near zero. Therefore, the key takeaway from Figures 3.7, 3.8, and 3.9 is that the differences in wave data from different sources were propagated to the vessel's responses, but the effect of these differences varied with speed and heading.

Despite the notable peak period and bandwidth differences (recall that the significant wave height, mean period, and mean direction were identical) between the measured and Bretschneider reconstructed buoy data shown in Figure 3.6 for Location B, the resultant responses of the two datasets were much more similar to each other than to the CMEMS and NWW3 data, as shown in the bias plots of Figure

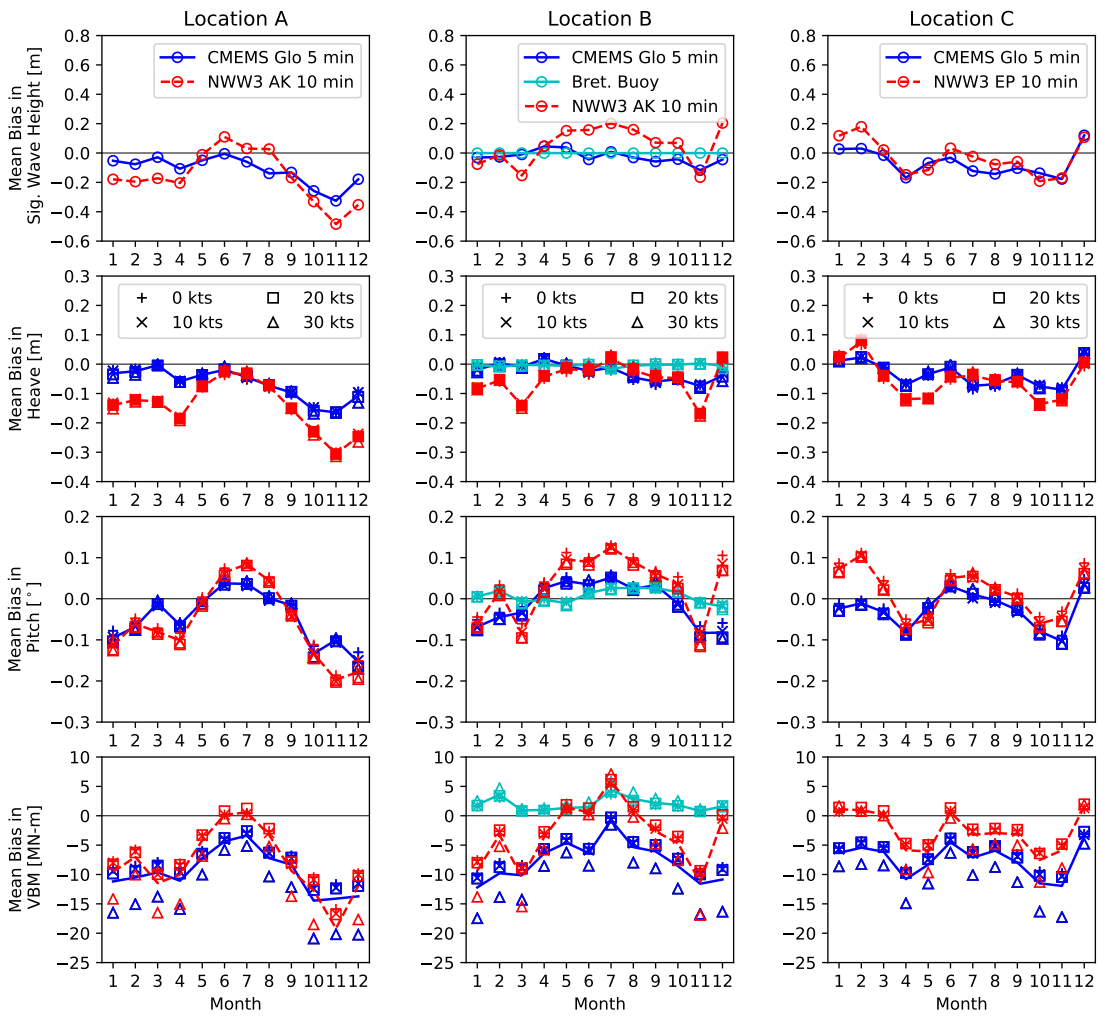

Figure 3.9: Monthly mean bias in significant wave height and the significant amplitudes of the heave, pitch, and VBM responses at Locations A, B, and C. The four speeds are plotted as different markers with a line showing the monthly mean bias over all speeds. Location B also includes the results of the Bretschneider reconstruction of the buoy data.

3.9. This similarity between buoy data with different spectral shapes implies that significant wave height and wave direction may have been more influential for response prediction than peak period and bandwidth, which are more directly linked to spectral shape than the former two parameters. Although wave peak period and wave bandwidth may not have been excessively detrimental to the predicted motions and VBM responses examined here, these parameters may have still had a notable impact in the calculation of fatigue damage, which is discussed further in Section 3.2.3.

To examine the propagation of bandwidth from wave spectra to the resultant VBM responses, Figure 3.10 shows the direction-averaged monthly wave and VBM bandwidth bias values as a function of vessel speed in addition to the wave bandwidth bias plots from Figure 3.6 for comparison. Although the plots of Figure 3.10 are direction-averaged despite bandwidth being heavily dependent on forward speed and direction through the the encounter frequency, Figure 3.10 does show that bandwidth differences exist between wave models, which directly influence fatigue damage as shown in Equation 2.12. The second row of Figure 3.10 may be the most informative because it was not influenced by forward speed. Comparing the wave bandwidth of the first row with the VBM response bandwidth of the second row, it appears that the VBM RAOs may have acted as a filter on the wave model spectra, which led to better agreement with the measured wave buoy data despite wave bandwidth differences. For the three rows with forward speed, it seems that the bandwidths of CMEMS and NWW3 were in better agreement than their respective wave datasets; however, for Location B, the Bretschneider reconstruction of the buoy data remained different than CMEMS, NWW3, and the measured buoy data.

3.2.3 Fatigue Damage Comparisons

The direction-averaged cumulative fatigue damage at Locations A, B, and C and the nearby buoys are shown in Figure 3.11. For each curve, the 13 plotted values include an assumed damage value of 0.0 at the beginning of the year and 12 values corresponding to the cumulative damage at the end of each month (including all prior months' damage). The horizontal black lines at $D = 0.06$ were included for reference to the ordinate scale between speeds. In Figure 3.11, it is evident that the differences between wave data sources led to notable differences in cumulative fatigue damage after just one year. These differences at the end of the one-year period are quantified in Table 3.3, which gives the percentage difference of direction-averaged

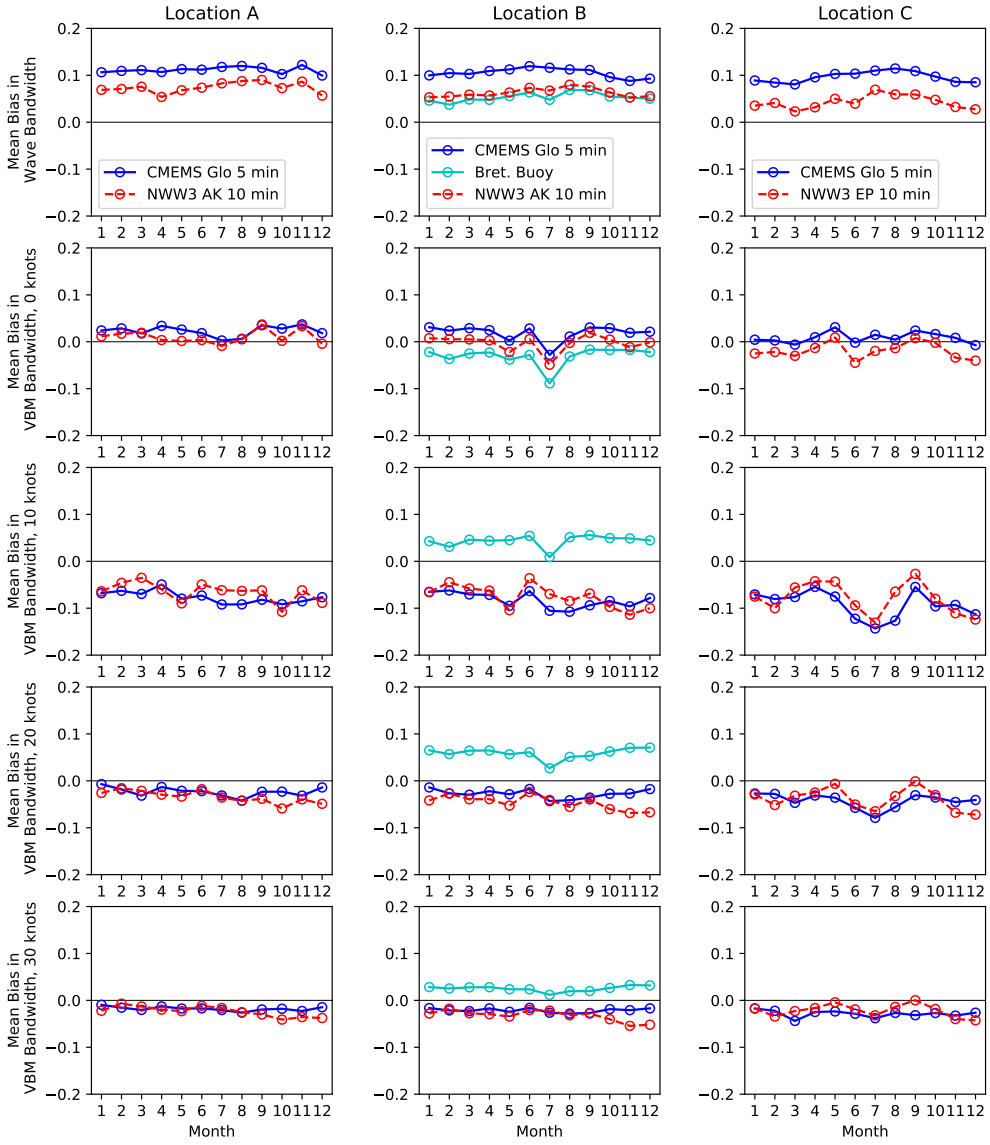

Figure 3.10: Monthly mean bias in wave and VBM bandwidth for each speed at Locations A, B, and C. Location B also includes the results of the Bretschneider reconstruction of the buoy data.

fatigue damage from each wave data source relative to the direction-averaged damage estimated using the measured NDBC buoy data (i.e., $100 \times (M_i - B_i)/B_i$ %).

In the first columns of Table 3.3 and Figure 3.11 corresponding to Location A, the relative percentage differences of CMEMS and NWW3 show that both wave model datasets yielded notably less damage than the measured buoy data, although the magnitudes of these differences were slightly greater for CMEMS than NWW3 at all

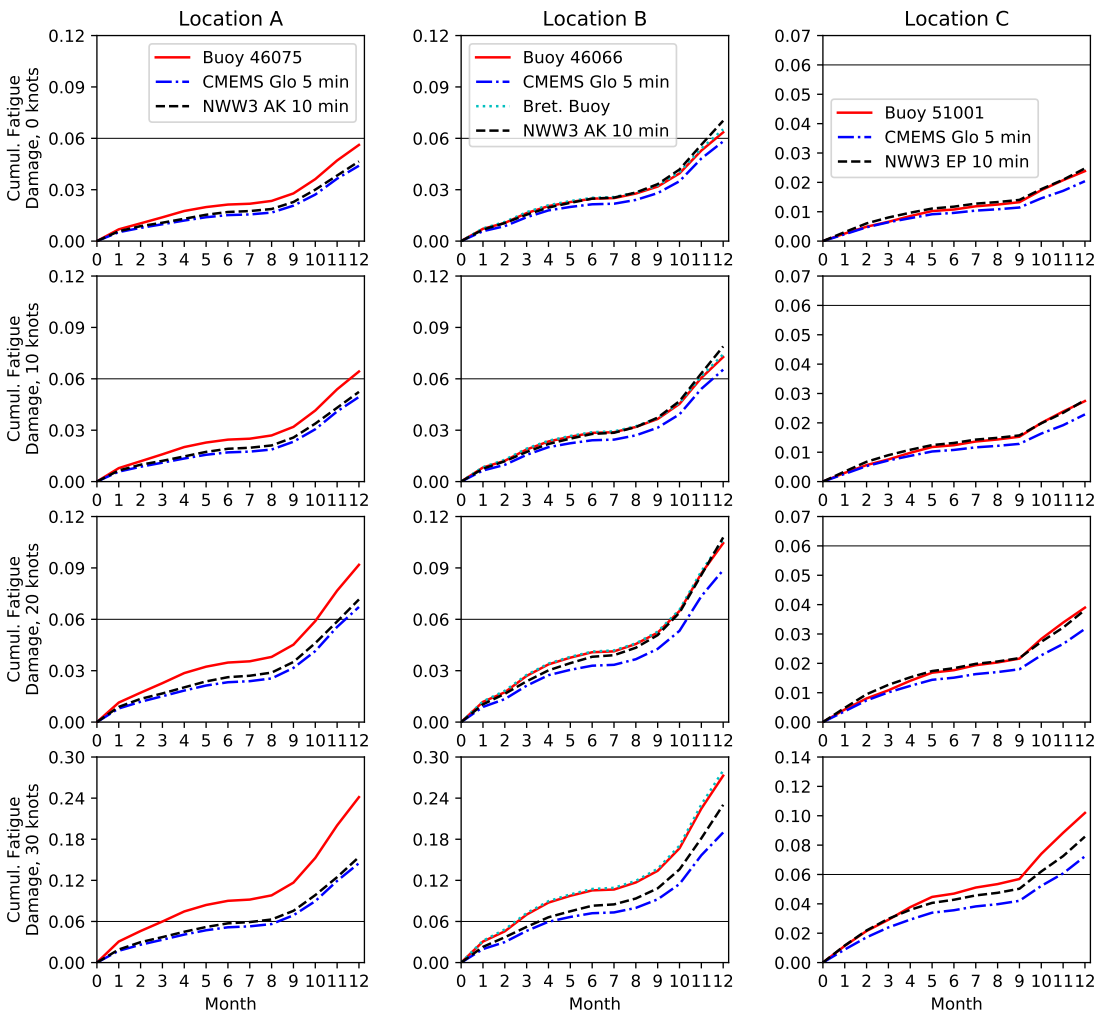

Figure 3.11: Direction-averaged cumulative fatigue damage at the end of each month for all four speeds at Locations A, B, and C. Note the change in $y-$axis scales between plots. Horizontal lines are included at D = 0.06 for reference. Location B also includes the results of the Bretschneider reconstruction of the buoy data.

four speeds. In contrast, for Locations B and C and speeds of 0, 10, and 20 knots, the relative percentage differences of CMEMS and NWW3 were separated by approximately 16-20%, with the CMEMS predicted damage typically less than the measured buoy, and NWW3 predicted damage greater than the buoy. For 30 knots at Locations B and C, while both CMEMS and NWW3 yielded less damage than the measured buoy data, CMEMS percentage difference magnitudes were nearly double that of NWW3. In Figure 3.9, the effect of speed on response bias was only notable for VBM

Table 3.3: Percentage differences in direction-averaged fatigue damage at the end of the year relative to the direction-averaged damage predicted using the measured NDBC buoy spectra (i.e., $100 \times (M_i - B_i)/B_i$ %).

Speed	Source	Location A	Location B	Location C
0 knots	CMEMS	-21.7 %	-8.4 %	-14.4 %
	NWW3	-17.1 %	10.7 %	3.9 %
	Bret. Buoy	–	2.8 %	–
10 knots	CMEMS	-23.1 %	-10.1 %	-16.5 %
	NWW3	-18.5 %	8.5 %	1.3 %
	Bret. Buoy	–	2.2 %	–
20 knots	CMEMS	-26.8 %	-15.0 %	-18.6 %
	NWW3	-22.0 %	3.2 %	-2.1 %
	Bret. Buoy	–	2.0 %	–
30 knots	CMEMS	-39.9 %	-30.4 %	-29.0 %
	NWW3	-36.2 %	-15.6 %	-15.9 %
	Bret. Buoy	–	2.8 %	–

responses at 30 knots, and the effect was more pronounced for CMEMS than NWW3, especially at Locations B an C. These speed effects are evident in the final rows of Table 3.3 and Figure 3.11, which show significantly more damage resulting from the buoy data than the CMEMS and NWW3 data at 30 knots. Note that given the length of this study and the harsh conditions encountered, some of the simulations performed (e.g., 30 knots in a severe sea state) were not entirely realistic; however, these results still provide insight into the significance of which wave hindcast dataset is selected for structural health monitoring. Furthermore, as mentioned in Section 3.1.6, while direction averaging of fatigue damage was necessary for tractability, less significant VBM responses and fewer loading cycles resulting from predominant beams seas may have detracted from the observed magnitude of fatigue differences in predominant head seas. On the contrary, the structural fatigue life assessment discussed by *Stambaugh et al.* (2014) for a US Coast Guard Cutter noted that vessel's often practice heavy weather avoidance whenever possible. Although infrequent, *Stambaugh et al.* (2014) showed that given a critical mission (e.g., search and rescue), traveling at undesirable headings and speeds in severe sea states may be unavoidable and result in

sudden large increases in fatigue damage. The simulations performed in this work considered eight evenly-spaced headings such that the likelihood of each heading was equivalent, making it possible that a higher occurrence of fatigue damage values from undesirable headings may have skewed the direction-averaged data to more significant damage values. Nevertheless, if the differences shown in Table 3.3 and Figure 3.11 were to continue accumulating over multiple years, the estimated fatigue damages would become drastically different in a fraction of the vessel's operating life.

As shown for Location B in Table 3.3 and Figure 3.11, the Bretschneider reconstruction of the buoy data yielded nearly identical fatigue damage results to the measured buoy data. As mentioned in Section 3.2.2, this similarity may imply that significant wave height and wave direction (identical between buoy datasets) were more influential than peak period and bandwidth, which are more directly related to spectral shape. However, this result may have been specific to the selected location, and it is clear through Equation 2.12 that bandwidth and period (zero up-crossing period, specifically) do have a direct impact on fatigue damage.

To more explicitly examine the effects of a change in significant wave height or peak period on the ship's VBM responses and fatigue damage accumulation rate, a brief demonstration was performed in head seas at 20 knots. In four different sea states (SS), Table 3.4 shows the percentage change in VBM significant amplitude, ζ_M, caused by a $\pm 10\%$ change in significant wave height (H_s) or peak period (T_p). The values of H_s and T_p were provided in *Bales* (1983) for open ocean in the North Pacific and used to construct a Bretschneider spectrum. The VBM significant amplitudes in the fourth column of Table 3.4 are the values calculated without a $\pm 10\%$ change in H_s or T_p. As shown in Table 3.4, a $\pm 10\%$ change in significant wave height yielded an identical percentage change in VBM response, which was expected given the linear response prediction approach of the initial twin framework (i.e., changing H_s just scales the the energy of the idealized wave spectrum before it is multiplied by the

Table 3.4: Percentage changes in VBM significant amplitude, ζ_M, resulting from a $\pm 10\%$ change in significant wave height, H_s, or peak period, T_p, for select sea states (SS) in head seas at 20 knots. The H_s and T_p values were provided in *Bales* (1983) for open ocean in the North Pacific.

SS	H_s [m]	T_p [s]	ζ_M [MNm]	$H_s + 10\%$	$H_s - 10\%$	$T_p + 10\%$	$T_p - 10\%$
3	0.88	7.5	22.1	+10.0%	−10.0%	+21.6%	−26.8%
4	1.88	8.8	62.8	+10.0%	−10.0%	+ 8.1%	−15.1%
5	3.25	9.7	117.5	+10.0%	−10.0%	+ 2.2%	− 8.4%
6	5.00	13.8	160.7	+10.0%	−10.0%	− 9.8%	+ 9.0%

square of the RAO magnitude in Equation 2.4). In contrast, a $\pm 10\%$ change in peak period showed potential to yield higher magnitude percentage differences in VBM significant amplitude (e.g., SS3 and SS4 in Table 3.4). Unlike significant wave height, a change in peak period means a shift in the frequency alignment between the wave spectrum and the RAO.

The corresponding percentage changes in fatigue damage due to changes in significant wave height and peak period are given in Table 3.5. The fatigue rates in the second column of Table 3.5 are the values calculated without a $\pm 10\%$ change in H_s or T_p. The magnitudes of the fatigue damage rate fluctuations due to a $\pm 10\%$ change in significant wave height demonstrate the amplification of VBM differences in the fatigue damage calculation of Equation 2.12, which is a higher-order function of stress. As shown in Table 3.5, a $\pm 10\%$ change in peak period has potential to cause even greater fatigue damage differences than a change in significant wave height. In addition to a change in VBM response, a shift in peak period also effects the number

Table 3.5: Percentage changes in fatigue damage rate resulting from a $\pm 10\%$ change in significant wave height, H_s, or peak period, T_p, for select sea states (SS) in head seas at 20 knots. The sea state definitions are provided in Table 3.4.

SS	Damage Rate $\times 10^6$ [- /hour]	$H_s + 10\%$	$H_s - 10\%$	$T_p + 10\%$	$T_p - 10\%$
3	0.4	+33.1%	−27.1%	+59.1%	−50.6%
4	7.2	+33.1%	−27.1%	+18.3%	−32.1%
5	44.0	+33.1%	−27.1%	+ 1.4%	−17.4%
6	96.7	+33.1%	−27.1%	−28.7%	+34.1%

of loading cycles experienced by the vessel. Although this demonstration was limited to head seas at 20 knots in four select sea states, it showed the transformation of seemingly benign significant wave height and peak period differences into notable differences in VBM response, which yielded potentially drastic differences in fatigue damage rate.

Despite the simplicity of the frequency-domain response prediction and fatigue damage estimation procedures of the initial twin framework, based on the wave data comparisons made in Section 3.2.1, it is expected that similar discrepancies between wave data sources would arise for higher fidelity hydro-structural models. The relative differences presented in this chapter show that selection of a wave data source can heavily influence the estimation of cumulative fatigue damage.

3.3 Conclusions

In this study presented in *Schirmann et al.* (2019b) and *Schirmann et al.* (2020a), at three different locations, hindcast wave data from multiple sources and the resultant heave, pitch, and VBM responses and fatigue damage estimates calculated for 32 speed and heading combinations were compared in three-hour increments for the entire year of 2017. These wave data included multidirectional hindcast model data from CMEMS and NWW3 in addition to measured wave spectra from NDBC buoys. Differences between these three wave data sources propagated to the vessel's heave, pitch, and VBM responses at all three locations. The VBM response differences were then magnified in the initial twin framework's calculation of fatigue damage, a higher-order function of stress, causing notable disagreement in damage estimates after just one year. At Location A, for each speed, the CMEMS and NWW3 data sources yielded similar fatigue damage accumulations at the end of the year, but this damage was significantly less than the damage estimated using measured buoy data. At Locations B and C and select speeds, CMEMS and NWW3 produced significant

differences in fatigue damage when compared to each other and the measured buoy. If these differences accrued over multiple years, the estimated damage for a vessel may vary drastically depending on the wave data employed in just a fraction of the vessel's service life.

As mentioned previously, the goal of this study was not to comment on the validity of these wave data sources or determine a hierarchy, but rather to demonstrate the significance of wave data source selection for response forecasting or hindcasting and digital structural health monitoring. Although the relationships between different wave model and buoy data sources examined here were potentially dependent on geography, this study demonstrated the challenges associated with wave data source selection for time-and-place specific response predictions and, in particular, fatigue damage estimates using wave data alone. These challenges are more significant now than ever before as the marine industry pushes towards digital vessel monitoring and eventually, autonomous vessels, both of which remove human judgment from real-time decisions. While higher fidelity seakeeping and hydro-structural models could be employed, the propagation of wave data differences to response predictions and fatigue damage estimates seen here suggests that the quality of results would be limited by the accuracy of the input wave data used. Despite the readily accessible wave data and advanced state of oceanographic models, through calculation of over three million responses, this investigation suggested that more study on wave data sources and their selection is required for reliable implementation of vessel response prediction and through-life fatigue analyses using wave data alone.

CHAPTER IV

Preliminary Motion and Vertical Bending Moment Correction Models

Armed with the insights from the initial, "forward-only" twin demonstration of Chapter II and the wave data source selection study of Chapter III, exploration of fusion approaches for development of reflective twin concepts could begin. As mentioned in Chapter I, the rapidly expanding field of data science has developed many ML tools that have extraordinary potential within the marine industry for improved asset management capabilities. To incorporate these ML algorithms and form a solid foundation for future development of a usable surface platform digital twin, the knowledge gap was first addressed using relatively simple data-driven models. As a starting point, the goal of this work was to compare the initial twin's response predictions, generated using unidirectional wave data, to full-scale data measured aboard a ship and apply two correction techniques to the predicted values: a linear least-squares (LS) approach and an artificial neural network (NN). The following investigation of these correction approaches for research vessel heave and pitch prediction was originally presented in *Schirmann et al.* (2019a). Also included in this chapter, these approaches were tested for correction of VBM values from a United States Coast Guard Cutter, which was originally discussed in *Hageman et al.* (2019) and *MARIN* (2019). The successes of these correction models, based on unidirectional wave data,

informed the application of more intensive data-driven models, based on multidirectional wave data, in subsequent chapters.

A number of studies have been performed that employed ML algorithms for vessel motion prediction, many of which employed different forms of NNs (e.g., fully-connected NNs, recurrent NNs, convolutional NNs). The successes of these studies demonstrate that data-driven models have merit for motion prediction applications. The following examples from the literature were presented in *Schirmann et al.* (2021). *Bremer* (2018) tested linear regression and NNs for sway, heave, and roll motion prediction in regular wave, beam-seas, model-scale experiments. Using just wave period and height as input variables, the results highlighted the ability of NNs to accurately predict vessel motions and model highly nonlinear phenomena (such as roll), a distinct advantage over linear regression. Using roll angle and roll velocity as input, *Xing and McCue* (2010) demonstrated the use of NNs to fit the parameters of two different nonlinear models for roll motion. *Li et al.* (2017) demonstrated the use of nonlinear autoregressive exogenous (NARX) networks for predicting pitch and roll angle time series, as well as several other ship motion characteristics. Experiments were performed using online, offline, and hybrid (combined online and offline) learning approaches, and results were discussed for single- and multi-timestep ahead prediction. A discussion of trade-offs between NNs and other machine learning approaches was also included, which ultimately selected NNs for vessel motion prediction. *Khan et al.* (2005) demonstrated the efficacy of NNs for accurate prediction of roll angles up to 10 seconds in the future for ship system deployment. *De Masi et al.* (2011) demonstrated radial basis function NNs (RBFNNs) for vessel heave time series predictions using prior heave time series data as input. Similarly, *Huang and Zou* (2016) compared RBFNNs with two other types of NNs for short-term pitch prediction. *Deng et al.* (2020) successfully demonstrated the use of long short-term memory (LSTM) networks, a form of recurrent NN, to predict semi-submersible platform heave, pitch,

and roll time-series due to irregular waves from a single direction, where the model inputs were wave elevations. Comparisons were made to motion predictions generated using experimentally determined and physics-based numerical model RAOs, which showed that the NN approach outperformed both RAO-based methods. Although the implementations differ significantly, the work discussed in this chapter similarly draws comparisons between the initial-twin framework's RAO-based predictions and data-driven model predictions. Nevertheless, the work performed in this dissertation, particularly in Chapter V, extends well beyond these comparisons and directly investigates the incorporation of PBMPs into data-driven models.

Numerous NN examples exist in the marine context beyond motion predictions, showing their prominence in the field. The following examples were also presented in *Schirmann et al.* (2021). *Jiang et al.* (2020) employed an NN within a submarine maneuvering model, and *Najafi et al.* (2018) used NNs to predict hydrofoil supported catamaran resistance, effective power, trim, and sinkage based on model-scale experiment data. NNs have also been applied to analyze the operational wave environment using ship motions. For example, *Duz et al.* (2019) utilized an NN that combined convolution layers with an LSTM network to predict wave parameters using ship motion time series. *Kawai et al.* (2021) used ship motion and structural response spectra as input to a convolutional NN that predicted wave parameters. NNs have also been applied in the context of oceanography for efficient wave modeling. For example, *Wu et al.* (2020) trained an NN using the environmental variables and boundary conditions typically used as input for computationally expensive wave models to efficiently generate multi-step wave forecasts. For more examples of NNs in marine applications, please see *Gougoulidis* (2008).

Other machine learning approaches have also been applied to hydrodynamic data in the past. For example, *Weymouth and Yue* (2013) demonstrated physics-based learning models for multiple applications, including seakeeping analyses using ex-

perimental data and prediction of ship bow wave behavior using computational fluid dynamics results. Nevertheless, published applications of machine learning algorithms using *real-world* data from operational, full-scale ships are limited. The two studies outlined in this chapter were first-step experiments with data-driven models trained using real-world data for reflective twins.

Full-scale motion measurements were acquired from an October and November 2013 cruise of the *R/V Knorr*, an 85 m oceanographic research vessel, from Nuuk, Greenland to Woods Hole, Massachusetts. Utilizing speed, heading, and location information from the ship in conjunction with unidirectional NWW3 wave hindcast data (*NOAA National Centers for Environmental Prediction.*, 2019), the initial twin framework was used to predict the vessel's motions during the cruise. The predicted and measured motions of the vessel were compared, and the LS and NN models were used to correct the predicted motions such that they more closely matched the measured values. The effectiveness of each correction model and the influence of different dataset partitioning approaches on correction model output were investigated.

To further test these correction approaches, this time in a structural context, vertical bending moment (VBM) data from the *Unites States Coast Guard Cutter (USCGC) Bertholf* were acquired through the VALID II Joint Industry Project (JIP), led by Maritime Research Institute Netherlands (MARIN). Although the dataset was notably larger than that from the *R/V Knorr*, it was not without limitations. Due to the controlled nature of the dataset, a set of RAOS, VBM measurements, speed, and relevant unidirectional wave data were provided directly by MARIN; location and heading information were not made available (the wave data were provided with directions relative to the ship heading). Once again, the goal was to correct the initial twin framework's VBM predictions such that they matched the measured responses more closely. The performances of the LS and NN correction methods were evaluated using a test dataset, including a comparison of results generated using two different data

filtering approaches. The knowledge gained from the two experiments outlined in this chapter formed a foundation for implementation of more complex data-driven models using significantly larger datasets (with fewer limitations) in subsequent chapters of this dissertation.

4.1 Methodology

This section details the methodology behind the LS and NN approaches primarily in the context of heave and pitch prediction correction for the *R/V Knorr* as originally presented in *Schirmann et al.* (2019a). The adaptation of this same general methodology for the *USCGC Bertholf* VBM dataset, originally discussed in *Hageman et al.* (2019) and *MARIN* (2019), is detailed at the end of this section.

A detailed description of the October and November 2013 oceanographic research cruise of the *R/V Knorr* is given in *Blomquist et al.* (2013), which is publicly available along with the data analyzed in the following discussion at `ftp://ftp1.esrl.noaa.gov/psd3/cruises/HIWINGS_2013`. Figure 4.1 shows a map of the cruise portions considered in this work. The data recorded during this cruise included location, heading, speed, and the motions of the vessel at sampling rates of up to 10 Hz throughout the duration of the voyage. Using time and location information, relevant wave data were acquired from the NOAA WAVEWATCH III (NWW3) Production Hindcast (*NOAA National Centers for Environmental Prediction.*, 2019). Although available through the same data source, the unidirectional NWW3 data employed here differed (in file format and wave parameters) from the multidirectional data used in the wave data source selection study of Chapter III. The unidirectional NWW3 wave data were provided at three-hour intervals. As such, the available full-scale data were analyzed in three-hour time windows, with the mean time of each window corresponding to a wave data interval time. The vessel's mean location, speed, and heading, as well as the standard deviations of measured heave (σ_{3M}) and pitch (σ_{5M}),

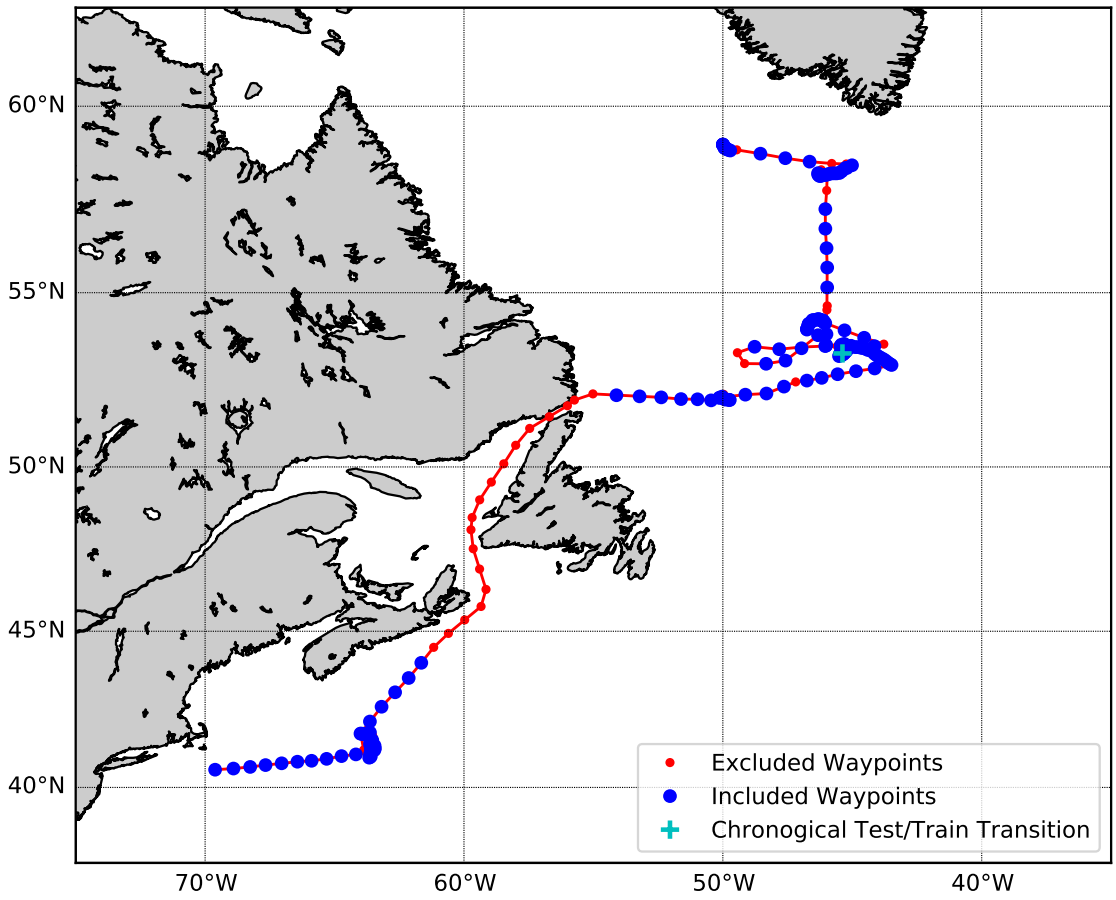

Figure 4.1: A map showing the analyzed waypoints of an October and November 2013 oceanographic research cruise of the *R/V Knorr*. Each waypoint is representative of a three-hour interval, and the waypoints included and excluded from further analysis are indicated.

were designated as representative values for each time window. Note that exclusion of time intervals with non-stationary heading and speed, which was performed to increase the validity of stationary assumptions, is discussed in Section 4.1.2. Also note that some geographically sheltered waypoints of the cruise shown in Figure 4.1 were excluded from further analysis. The data points excluded from and included in further analysis are indicated in Figure 4.1. Different methods of partitioning the waypoints into test and train datasets for the correction models are presented in Section 4.1.3, which includes a discussion of the chronological test/train transition point of Figure 4.1.

Heave and pitch were measured aboard the vessel by an *Applanix POS MV* system. Note that in this study, heave refers to the vertical displacement at the *POS MV* system, which was slightly forward of midships and off of the centerline of the vessel (*Doherty and Berni Associates Inc.*, 2004); therefore, the frequency-domain response predictions of heave were calculated at the location of this motion measurement device rather than the vessel's center of gravity. This subtlety means that the heave motions discussed in the following analysis also included the impact of pitch and roll motions.

4.1.1 Response Predictions

This work utilized `SHIPMO.BM` to calculate the *R/V Knorr*'s RAOs for the initial twin framework. RAOs were generated for speeds of 0 to 13 knots in increments of 1 knot and relative wave headings of 0 to 355° in increments of 5°. As mentioned previously, unlike the initial twin demonstration of Chapter II and wave data source selection study of Chapter III, the NWW3 wave data used in this work were unidirectional; therefore, the conditions for a given three-hour time interval were defined by a single wave partition characterized by a significant wave height (H_s), peak period (T_p), and mean direction. The choice to use unidirectional wave data was made to simplify these preliminary investigations by narrowing the field of input variables for the initial correction models. The use of multidirectional wave data for data-driven approaches is discussed in subsequent chapters.

A Bretschneider spectrum, $S_B^+(\omega_0)$, was constructed for each time window using H_s and T_p. In contrast to the work of Chapters II and III, because unidirectional wave data were used, a cosine-squared spreading function of ±90° in increments of 15° was applied to the wave data to simulate short-crested seas, as discussed in *Lloyd* (1989). The one-sided pseudo-response spectrum of the vessel for the j^{th} degree of freedom, $S_j^+(\omega_0)$, where $j = 3$ for heave and $j = 5$ for pitch, was calculated using Equation

4.1, which is a modified version of Equation 2.4 that incorporates wave spreading.

$$S_j^+(\omega_0) = \frac{2}{\pi} \int_{-\pi/2}^{\pi/2} S_B^+(\omega_0) |RAO_j(\omega_0, \theta + \mu)|^2 cos^2(\mu) d\mu \tag{4.1}$$

In Equation 4.1, θ is the heading of the ship relative to the incoming waves during a given three-hour interval and μ is the spreading angle. Note that $RAO_j(\omega_0, \theta+\mu)$ was also a function of speed, and the `SHIPMO.BM` RAO with speed and relative heading closest to U and $\theta+\mu$, respectively, was used as $RAO_j(\omega_0, \theta+\mu)$ for each 15° ($\frac{\pi}{12}$ rad) discrete integration step of Equation 4.1. The vessel response calculation procedures of the initial twin framework following Equation 2.4 remained unchanged.

4.1.2 Uncorrected Predictions and Data Filtering

The frequency-domain response prediction approach of the initial twin framework required the assumption of statistical stationarity during each three-hour interval. To increase the validity of this assumption, heading and speed standard deviation filters were applied to remove data points from time windows where conditions were significantly non-stationary. These filters were defined by thresholds, where data points with speed or heading standard deviations greater than these thresholds were excluded from further analysis. To determine appropriate filter thresholds, the cumulative mass functions of the dataset's heading and speed standard deviations, shown in Figure 4.2, were analyzed. Using Figure 4.2, the speed standard deviation filter was set at 1 knot, and the heading standard deviation threshold was set at 22.5°. These thresholds are shown as vertical black lines in Figure 4.2. Note that data points with average speeds less than -0.5 knots were also excluded from further analysis because RAOs were only generated for non-negative speeds. As a reminder, all excluded data points are indicated in the map of Figure 4.1.

A comparison of measured and predicted heave (σ_{3P}) and pitch (σ_{5P}) standard

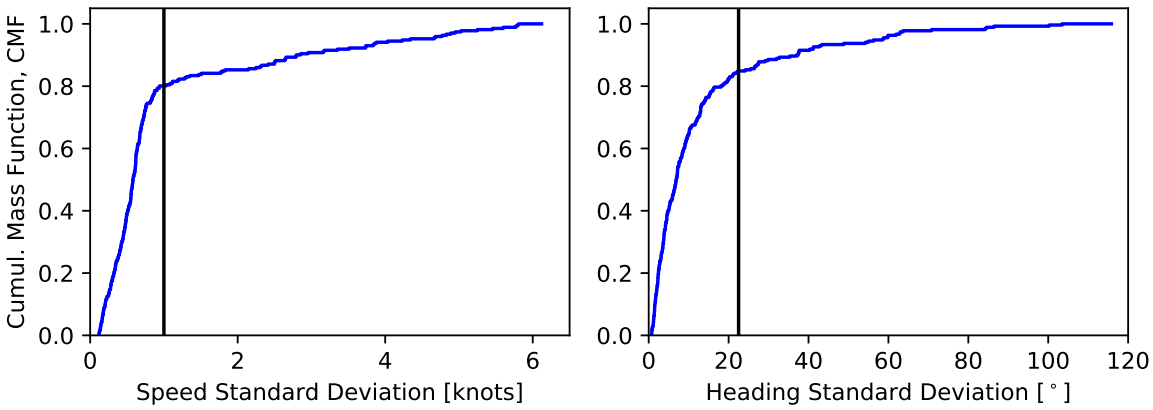

Figure 4.2: Cumulative mass functions (CMFs) of the speed and heading standard deviations for the unfiltered dataset. The vertical lines indicates the chosen filter thresholds.

deviations for the 271 total data points is given in Figure 4.3, which also shows the split between the 183 included and 88 excluded data points. Each point is representative of a three-hour time interval. The results shown in Figure 4.3 were calculated using the initial twin framework. These data are the uncorrected response predictions discussed in the remainder of this chapter for the *R/V Knorr*. In Figure 4.3, the black reference lines with slope 1.0 represent perfect agreement between predicted and measured values. As shown in Figure 4.3, the heave values were generally under-predicted. A slight under-prediction was also evident in the pitch data. Note that while exclusion of data points with relatively large speed and heading standard deviations was deemed necessary for the assumption of a stationary process, the excluded data points do not appear to be obvious outliers in Figure 4.3. This realization indicates that the error between predicted and measured variables may have been heavily influenced by H_s and T_p rather than U and θ alone, and that the assumption of linearity for the frequency-domain response calculations of the initial twin framework may have been a large cause of the witnessed under-predictions in heave and pitch.

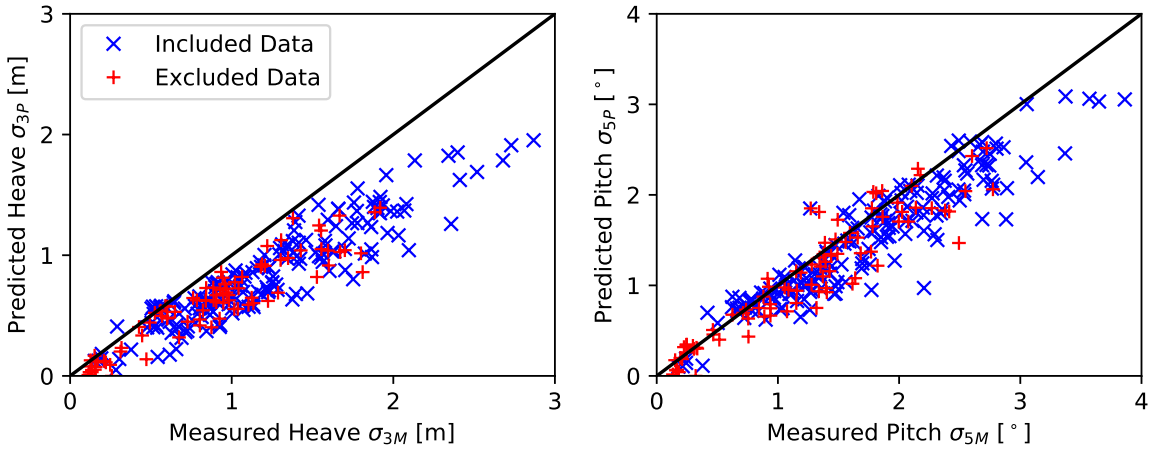

Figure 4.3: Predicted versus measured heave and pitch standard deviations, showing the 183 included data points and 88 excluded (filtered-out) data points. The black lines with slope 1.0 represent perfect agreement between predicted and measured values.

4.1.3 Data Partitioning

Two dataset partitioning methods were tested in this work with the *R/V Knorr* dataset. Both of these methods split the 183 filtered data points into two sets of approximately equal size (92 train and 91 test). Although it is common in ML applications to provide more data for training than testing (e.g., 80/20 train/test split), the limited size and diversity of this dataset encouraged the use of a larger test set for adequate model performance evaluation.

The first data partitioning method was to split the dataset chronologically into two sets, with the first as train data and the second as test data. The geographic transition point between test and train datasets is included in Figure 4.1. The second partitioning method was to randomly sort the points into train and test sets. For analyses using the random partitioning method, the dataset was re-partitioned between trials. A drawback of the chronological partitioning approach was that the duration of this journey was short relative to the lifetime of the vessel and confined to a small geographic region, so the experienced conditions (combinations of H_s, T_p, θ, U) were limited and, as such, there could be minimal overlap between the train

and test datasets. However, this chronological split was representative of a potential real-world application of machine learning corrections to predicted responses in which past observations are used as train data for future, unseen conditions. The random partitioning method was included in this work to demonstrate what a chronological split may yield if more diverse conditions were available to train the correction models. Figure 4.4 shows the chronological split between test and train datasets.

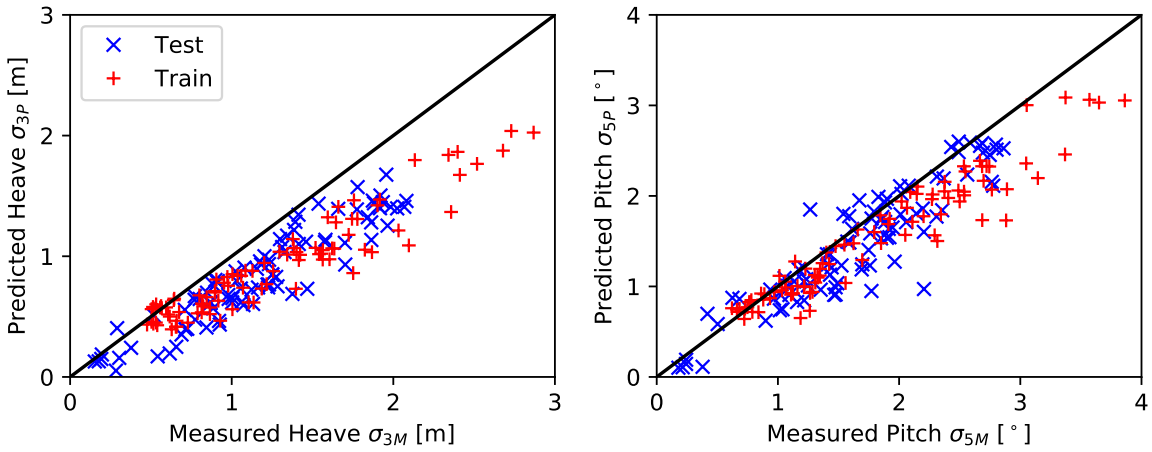

Figure 4.4: Predicted versus measured heave and pitch standard deviations, showing the test and train datasets partitioned chronologically. The black lines with slope 1.0 represent perfect agreement between predicted and measured values.

As shown in Figure 4.4, the measured value range of the train dataset extended to heave and pitch values greater than the test value range, and the test data range extended to lower measured values than the train data range. A larger and more diverse dataset from multiple cruises would likely have a less significant discrepancy between test and train data ranges. Probability also implies that in general, the discrepancy between randomly partitioned test and train data ranges would be less than the discrepancy shown in Figure 4.4 for the chronologically partitioned data. To further emphasize this discrepancy, the differences between test and train dataset input variables are compared in Figure 4.5, which shows the percent error of the predicted pitch standard deviations as a function of input variables. Equation 4.2 was used to calculate the percent error. While similar error plots for heave were

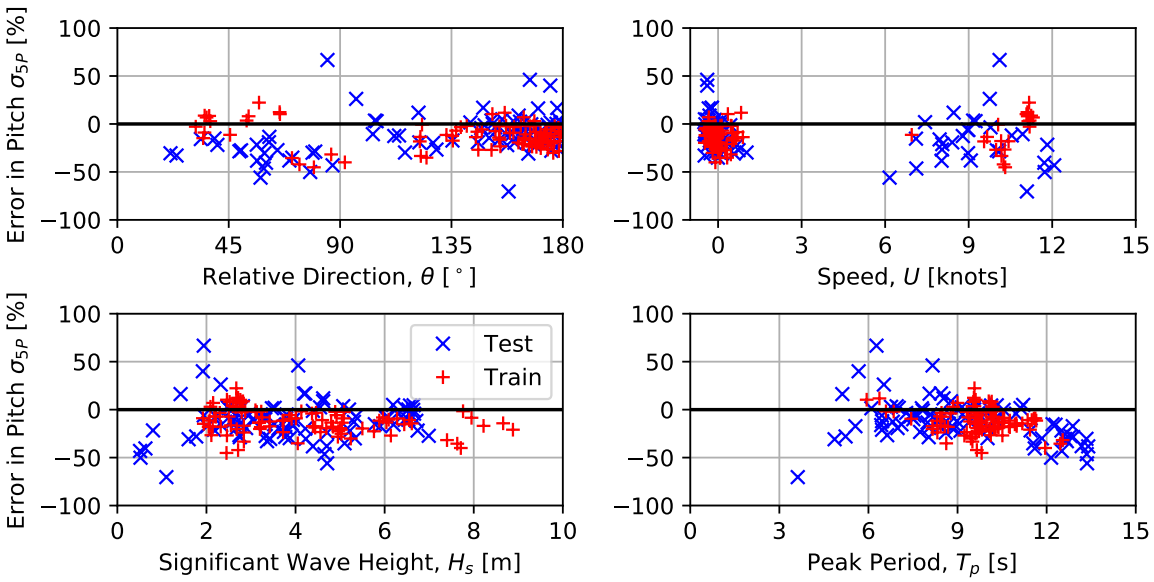

Figure 4.5: Error in predicted pitch standard deviations as a function of input variable, showing the test and train datasets partitioned chronologically.

excluded from this discussion, the percent error values were of the same order of magnitude as the pitch plots given in Figure 4.5.

$$\text{Error} = 100\left(\frac{\sigma_{jP} - \sigma_{jM}}{\sigma_{jM}}\right)\% \tag{4.2}$$

As shown in Figure 4.5, there are differences between the ranges of operation conditions experienced in the chronologically partitioned train and test datasets. Furthermore, it appears that the test dataset has more outliers than the train dataset. As mentioned previously, for trials using randomly partitioned data, probability suggests that more similar train and test input variable ranges would be expected than in the chronologically partitioned dataset of Figure 4.5.

4.1.4 Prediction Correction Methods

Two correction methods were applied to the predicted response standard deviations, σ_{jP}: a linear least-squares (LS) correction, and a neural network (NN) correction. In this work, a train data matrix, $\mathbf{X}_{Train}$, was constructed that contained the

predicted σ_{3P} and σ_{5P} of all train data, as well as the input variables (H_s, T_p, θ, and U) used by the initial twin framework to calculate these responses. More specifically, each of the 92 rows of $\mathbf{X}_{Train}$ corresponded to a single three-hour interval included in the train dataset, and the 6 columns of $\mathbf{X}_{Train}$ corresponded to H_s, T_p, θ, U, σ_{3P}, and σ_{5P}. Additional information such as vessel loading conditions, which were assumed to be constant between observations in this work, could be incorporated in future correction models if data beyond the design values are known. A matrix of test data, $\mathbf{X}_{Test}$, was also created with an analogous construction to that of $\mathbf{X}_{Train}$, with 91 rows and 6 columns. Train and test data matrices of measured heave and pitch standard deviations, $\mathbf{Y}_{Train}$ and $\mathbf{Y}_{Test}$, were also constructed. The 92 rows of $\mathbf{Y}_{Train}$ and the 91 rows of $\mathbf{Y}_{Test}$ corresponded to the rows of $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$, respectively, and each matrix's 2 columns corresponded to σ_{3M} and σ_{5M}, respectively. The 4 columns of $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$ corresponding to input data (H_s, T_p, θ, and U) were unit normalized by the maximum value found between both the train and test datasets; e.g. the H_s columns of $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$ were both normalized by the same value, which was the maximum H_s value seen in the entire dataset. The other two columns of $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$, corresponding to σ_{3P} and σ_{5P}, as well as the columns of $\mathbf{Y}_{Train}$ and $\mathbf{Y}_{Test}$, were unit normalized by the maximum measured heave and pitch values contained in $\mathbf{Y}_{Train}$ and $\mathbf{Y}_{Test}$. The normalized results of both correction models were multiplied by these normalization factors for analysis. Overviews of the linear LS and NN approaches used in this work are given below. For more information on these approaches, please see *Bishop* (2006) or *Russell and Norvig* (2010).

4.1.5 Linear Least-squares (LS) Correction

The linear LS problem formulation and solution matrix, $\mathbf{W}_{LS}$, are derived from the train dataset as shown in Equations 4.3 and 4.4. This solution minimizes the mean squared error (MSE) between measured and predicted train data values. Note

that in this formulation, a single column of ones is appended to both $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$ that acts upon a single row of bias terms (comprised of just two values, one for each degree of freedom) contained in $\mathbf{W}$ and $\mathbf{W}_{LS}$.

$$\mathbf{W}_{LS} = \arg\min_{\mathbf{W}} ||\mathbf{X}_{Train}\mathbf{W} - \mathbf{Y}_{Train}||_F^2 \tag{4.3}$$

$$\mathbf{W}_{LS} = [\mathbf{X}_{Train}^T\mathbf{X}_{Train}]^{-1}\mathbf{X}_{Train}^T\mathbf{Y}_{Train} \tag{4.4}$$

In Equation 4.3, $||\cdot||_F$ denotes the Frobenius norm. Using $\mathbf{W}_{LS}$, the matrix of LS corrected response predictions for the test dataset, $\mathbf{Y}_{LS,Test}$, was calculated using Equation 4.5.

$$\mathbf{Y}_{LS,Test} = \mathbf{X}_{Test}\mathbf{W}_{LS} \tag{4.5}$$

With an analogous construction to $\mathbf{Y}_{Test}$, the two columns of $\mathbf{Y}_{LS,Test}$ corresponded to the LS corrected heave and pitch response standard deviations, σ_{3LS} and σ_{5LS}, of the test dataset. Note that the LS approach can be used to model nonlinear phenomena (e.g., quadratic) via the input variables (e.g., inclusion of U and U^2) while still considered a linear model with respect to the learned weight parameters in $\mathbf{W}$; however, in this work, the LS approach was linear in both the input variable space and the weight parameters.

4.1.6 Neural Network (NN) Correction

All NN construction, training, and testing performed in this study utilized the `Flux` library in `Julia` (*Innes*, 2018). The NN correction models had input and output layers that used linear activation functions, while the intermediate, or hidden layer (HL) neurons used the LeakyReLU activation function, which is nonlinear. These nonlinear activation functions allowed the NN to model nonlinearities that could

not be captured by the LS correction. The input layer was composed of six neurons corresponding to the six columns of $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$. The output layer was composed of two neurons corresponding to the two columns of $\mathbf{Y}_{Train}$ and $\mathbf{Y}_{Test}$. Multiple architectures were tested for the HLs of the NN. The differences between architectures were the number of HLs, ranging from 1 to 5, and the number of neurons per HL, ranging from 3 to 15 in increments of 3 neurons, yielding 25 different candidate architectures. The Adam optimizer and MSE loss function were used to train the network with a batch size of 1. Note that in this work, the MSE between two matrices was the mean of individual column MSE values; therefore, both heave and pitch were considered equivalently when training and testing the NN. While different optimizers, loss functions, and HL activation functions were experimented with in this study, only the most successful selections were included here.

In this study, 20 total trials of data partitioning and NN construction and training were performed: 10 trials using chronological partitioning and 10 trials using random partitioning. Note that the train and test datasets were consistent for the 10 chronologically partitioned trials, but changed between each of the 10 randomly partitioned trials. In each of the 20 total trials, all 25 HL architectures were tested. For each architecture, the network was trained using the data contained in $\mathbf{X}_{Train}$ and $\mathbf{Y}_{Train}$ for a sufficient number of epochs, where a single epoch is defined as a pass through all training data points once. During this training process, the NN "learned" from the train data, and therefore evolved between each training epoch. During each training epoch, the current version of the NN was also applied to the test data contained in $\mathbf{X}_{Test}$, and the MSE between $\mathbf{Y}_{Test}$ and the resultant NN corrected test data matrix, $\mathbf{Y}_{NN,Test}$, was calculated. This test data MSE value characterized the effectiveness of the NN in its current state for test data correction. For each of the 25 tested architectures, the state of the NN model that yielded the lowest test data MSE was stored. The NN architecture and state that resulted in the lowest test data MSE was

selected as the representative model for each of the 20 trials. Note that this selection process, which relied upon the performance of the models when applied to the test dataset, means that the test dataset was actually more of a validation set than a true, fully-independent test set; however, given the very limited dataset size and the designation of this work as a preliminary exploration, this procedure was deemed most appropriate. Rigorous model selection processes using formal cross-validation and strict performance evaluations using fully-independent test sets are discussed in subsequent chapters. For each of the 20 trials performed in this study, the LS correction was also retrained and applied to each trial's test dataset for comparison.

4.1.7 Adaptations for *USCGC Bertholf* VBM Correction

As mentioned previously, the details of this of this study were originally discussed in *Hageman et al.* (2019) and *MARIN* (2019). The structural response data measured aboard the *USCGC Bertholf* and provided by MARIN were VBM standard deviations representative of 30-minute intervals dispersed intermittently between September 2010 and October 2012. Recall that due to the controlled nature of the dataset, the locations and true compass headings of the vessel were not disclosed in the provided dataset. As such, MARIN provided time-and-place specific unidirectional wave data corresponding to each 30-minute interval, characterized by a significant wave height, peak period, and peak direction relative to the vessel heading. These wave data were derived (by MARIN) from a ship-as-a-wave-buoy approach coupled with wave radar data for relative direction determination. Other provided information included the average vessel speed and the standard deviations of speed and heading recorded during each 30-minute interval. After removing all data with known errors or non-numeric values in the wave data or measured values, 5,842 data points remained, each representative of a 30-minute interval.

4.1.7.1 Response Predictions

VBM response amplitude operators (RAOs) were provided by MARIN for vessel speeds of 0, 5, 10, 15, 18, 21, and 28 knots and relative wave directions of 0 to 360° in 10° increments. These RAOs were generated using Bureau Veritas' software `HOMER` (*Malenica et al.*, 2013). As in the *R/V Knorr* implementation, VBM responses were predicted using the initial twin framework, which constructed a Bretschneider (ITTC two-parameter) spectrum; however, directional spreading was not applied in this experiment. No interpolation between speeds or relative directions was performed; the RAO with the closest speed and relative direction was selected for each time interval.

4.1.7.2 Data Filtering

As in the *R/V Knorr* heave and pitch prediction work, the frequency-domain analysis relied upon the assumption of stationarity during each 30-minute interval. As such, the 5,842 data points were filtered based on speed and heading standard deviations during each 30-minute interval. A 2.5 knot speed standard deviation filter was applied, where data points representative of a 30-minute interval with speed standard deviations below this threshold were kept. This dataset reduction yielded 5,007 data points. Two heading standard deviation filters were tested on this speed-filtered dataset. The first filter was a 45° heading standard deviation filter, which yielded a dataset with 4,361 points. The second filter applied to the speed-filtered dataset was a 5° heading standard deviation filter, which yielded a dataset with 2,964 points. Consistent with one partitioning method from the *R/V Knorr* study, the two filtered datasets were each partitioned randomly into two sets of approximately equal size: a train dataset and a test dataset. Unique to this VBM experiment, one final filter was applied to these resultant datasets, which was a ±100% error threshold between the measured and predicted values as calculated using Equation 4. This final reduction resulted in 4,054 (2,029 train and 2,025 test) points for the

45° heading filtered dataset and 2,732 (1,361 train and 1,371 test) points for the 5° heading filtered dataset. This error filter was applied to remove any remaining data points with erroneous values, possibly resulting from strain-gauge measurement noise, that could affect the training quality of the correction methods.

4.1.7.3 LS and NN Corrections

The main difference between the LS and NN model constructions used for VBM correction and those used for heave and pitch correction was the replacement of the heave and pitch standard deviation columns in $\mathbf{X}_{Train}$ and $\mathbf{X}_{Test}$ with a single column containing VBM standard deviations. Recall that these standard deviations were those predicted using the initial twin framework. Of course, the $\mathbf{Y}_{Train}$ and $\mathbf{Y}_{Test}$ matrices, which had two columns for heave and pitch, were also reduced to a single column for VBM correction. Unit normalization of the data was applied in an analogous manner to its use in the heave and pitch models. Although conducted in a less formal manner than for the heave and pitch models, multiple NN architectures were tested with various numbers of layers, activation functions, and neurons per layer. Once again, this selection process was performed using the test dataset for validation rather than rigorous performance evaluation. Many of the tested architectures yielded similar results to that of the final architecture discussed here. The final NN selected (for both the 5° and 45° heading filtered datasets) had five HLs with 60 neurons each, and an output layer with one neuron. The single output neuron corresponded to the corrected VBM standard deviation predictions. As in the heave and pitch correction models, the input and output layers used linear activation functions, the HLs used the LeakyReLU activation function, and the MSE loss function was employed for NN training. Once again, the `Flux` library in `Julia` was used to construct, train, and test the NN.

4.2 Results and Discussion

The performances of the LS and NN correction models for heave and pitch motions of the *R/V Knorr*, originally discussed in *Schirmann et al.* (2019a), and for VBM responses of the *USCGC Bertholf*, originally discussed in *Hageman et al.* (2019) and *MARIN* (2019), are detailed in this section. Due to the preliminary nature of these experiments and aforementioned dataset limitations, recall that the test results outlined in this section are actually validation results rather than true, independent performance evaluations. These experiments formed a foundation for rigorous model training and evaluation with larger and more accessible datasets in subsequent chapters of this dissertation.

4.2.1 *R/V Knorr* Motion Corrections

Table 4.1 provides the NN HL architecture that yielded the lowest test data MSE in each of the 20 total trials and was therefore selected as the representative model for the trial. Note the significant differences between the best-performing NN models from the different trials, which is discussed in detail at the end of this section. Also included in Table 4.1, two additional trials were performed for visual demonstration of the correction models: one using chronological partitioning, with results presented in Figure 4.6, and one using random partitioning, with results presented in Figure 4.7. These example trials are also discussed later in this section. While both the LS and NN correction methods incorporated heave and pitch information equivalently,

Table 4.1: NN HL architectures, defined as number of layers/neurons per layer, of the NNs that yielded the lowest MSE test loss in each of the 20 total trials. The two additional trial architectures presented in Figures 4.6 and 4.7 are also included.

Trial	1	2	3	4	5	6	7	8	9	10	Fig. 4.6/4.7
Chronological	2/9	5/12	1/6	1/6	5/12	1/9	4/12	4/9	2/15	3/12	4/9
Random	5/12	1/15	3/12	1/9	3/9	2/9	2/9	5/15	4/9	3/12	1/15

the following evaluations of correction effectiveness considered these two motions independently. For the 10 trials of each data partitioning method, the effectiveness of each correction model was quantified through comparison of a corrected test data MSE to the uncorrected test data MSE. More explicitly, the metric outlined in Equation 4.6 was used to quantify the change in test data MSE resulting from an applied correction model for a given trial. The smaller the value given by Equation 4.6, the greater the correction effectiveness.

$$\text{Percent of Uncorrected } \sigma_{jP} \text{ MSE} = 100 \left(\frac{\text{Corrected } \sigma_{jP} \text{ MSE}}{\text{Uncorrected } \sigma_{jP} \text{ MSE}} \right) \% \tag{4.6}$$

The effectiveness of the LS and NN correction models are characterized in Table 4.2 for both the chronological and random partitioning methods. The values of Table 4.2 were calculated using Equation 4.6. The Mean, Best, and Worst columns of Table 4.2 are representative statistics for all 10 trials performed using each data partitioning method. The Mean columns represent the average effectiveness of the correction method over the 10 trials for each partitioning method. The Best and Worst columns of Table 4.2 represent the minimum and maximum percent of uncorrected σ_{jP} MSE seen over the 10 trials, respectively. Note that each value in Table 4.2 was

Table 4.2: Percentages of uncorrected test data σ_{jP} MSE, calculated using Equation 4.6, resulting from the LS and NN corrections using both partitioning methods. The Mean, Best, and Worst columns correspond to the mean, minimum, and maximum values from the 10 trials of each partitioning method, respectively. The results shown in Figure 4.6 and Figure 4.7 are also included.

	Chronological Split				Random Split			
Correction	Mean	Best	Worst	Fig. 4.6	Mean	Best	Worst	Fig. 4.7
Heave LS	19.7	19.7	19.7	19.7	13.6	11.0	18.1	12.7
Heave NN	15.5	12.7	19.1	15.6	11.1	9.8	13.6	10.2
Pitch LS	80.1	80.1	80.1	80.1	45.5	39.4	52.5	41.9
Pitch NN	67.0	59.6	73.9	65.8	39.0	31.7	44.3	34.7

calculated using the uncorrected MSE of a specific trial. While the uncorrected MSE did not change between trials using chronological partitioning because the train and test datasets were consistent over all trials, the uncorrected MSE of the randomly partitioned dataset did change between trials because the data were re-partitioned for every trial.

In Table 4.2, the heave and pitch LS statistics for the chronologically partitioned data are equivalent across their respective rows because the LS correction method was deterministic; therefore, the results did not change between trials because the train and test datasets did not change. Despite the consistency in train and test datasets across the chronologically partitioned trials, the heave and pitch NN results still changed between trials due to the stochastic nature of the Adam optimizer used to train the network. As implied by the statistics of Table 4.2, for both the chronologically and randomly partitioned data, the NN correction outperformed the LS correction for heave and pitch in all 10 trials. The LS correction used in this work was strictly linear, which proved to be very effective at reducing test data MSE for both the chronologically and randomly partitioned datasets. The advantage of the NN correction model was that it could account for linear and nonlinear input data dependencies that caused discrepancies between predicted and measured motions. However, it should be noted that the differences between performances of the LS and NN corrections in terms of MSE reduction were small relative to their improvement upon the initial uncorrected MSE.

Figures 4.6 and 4.7 show the LS and NN corrected heave and pitch test datasets using the chronological and random partitioning methods, respectively. The upper plots of Figures 4.6 and 4.7 show the LS corrected test data overlaid on the uncorrected test data, and the lower plots show the NN corrected test data overlaid on both the LS corrected and uncorrected test data. As mentioned in Section 4.1.2, it is visually evident that the uncorrected test data heave and pitch standard deviations were

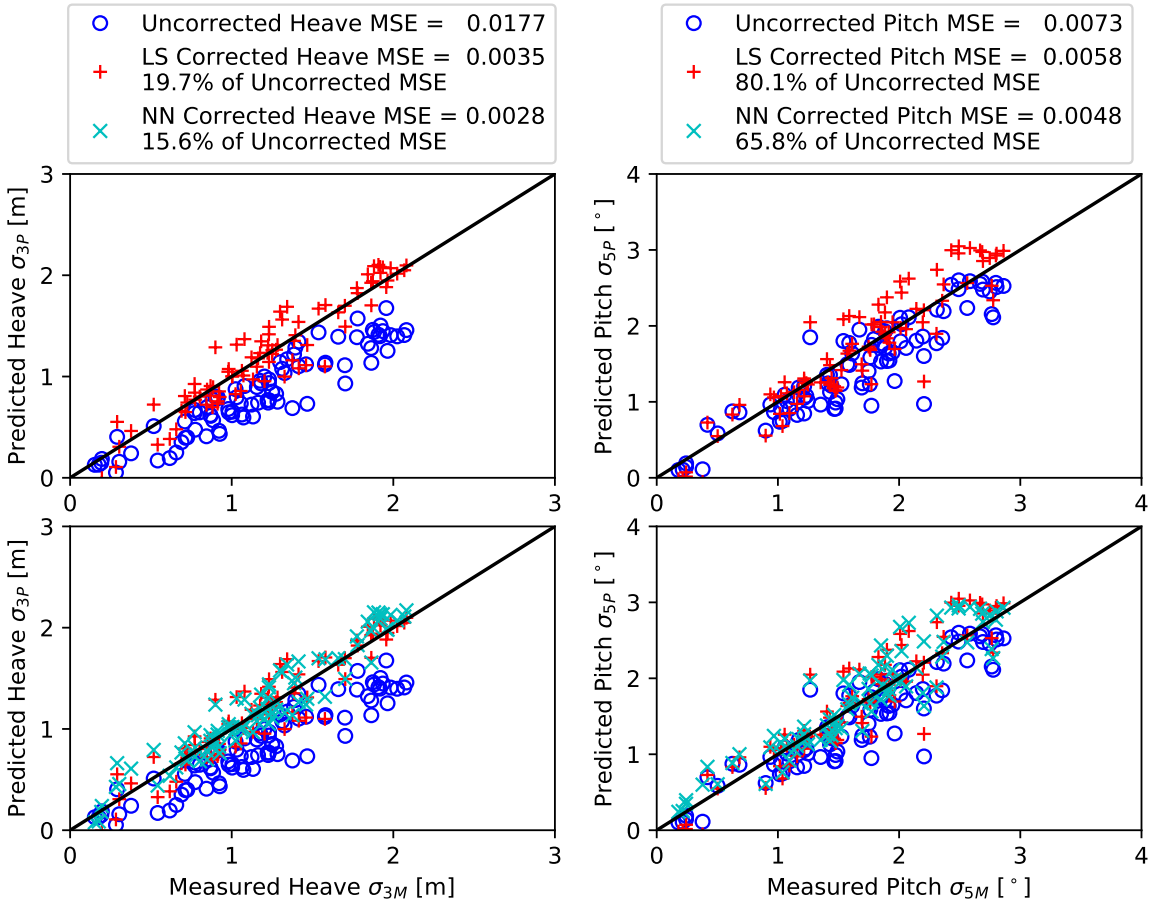

Figure 4.6: Predicted versus measured heave and pitch standard deviations for the chronologically partitioned uncorrected, LS corrected, and NN corrected test datasets. The black lines with slope 1.0 represent perfect agreement between predicted and measured values.

under-predicted relative to the measured values. In both Figure 4.6 and Figure 4.7, the LS correction effectively adjusted the test datasets such that the points more closely followed the black reference lines with slope 1.0, which represented perfect agreement between predicted and measured values. The effectiveness of the linear LS correction was confirmed by the reduction of MSE outlined in Table 4.2. Note that the heave MSE reduction is significantly greater than the pitch reduction because heave had a more significant under-prediction in the uncorrected test datasets.

As shown in Table 4.2, the 10 trials performed using random partitioning generally yielded more significant heave and pitch MSE reductions than the trials using

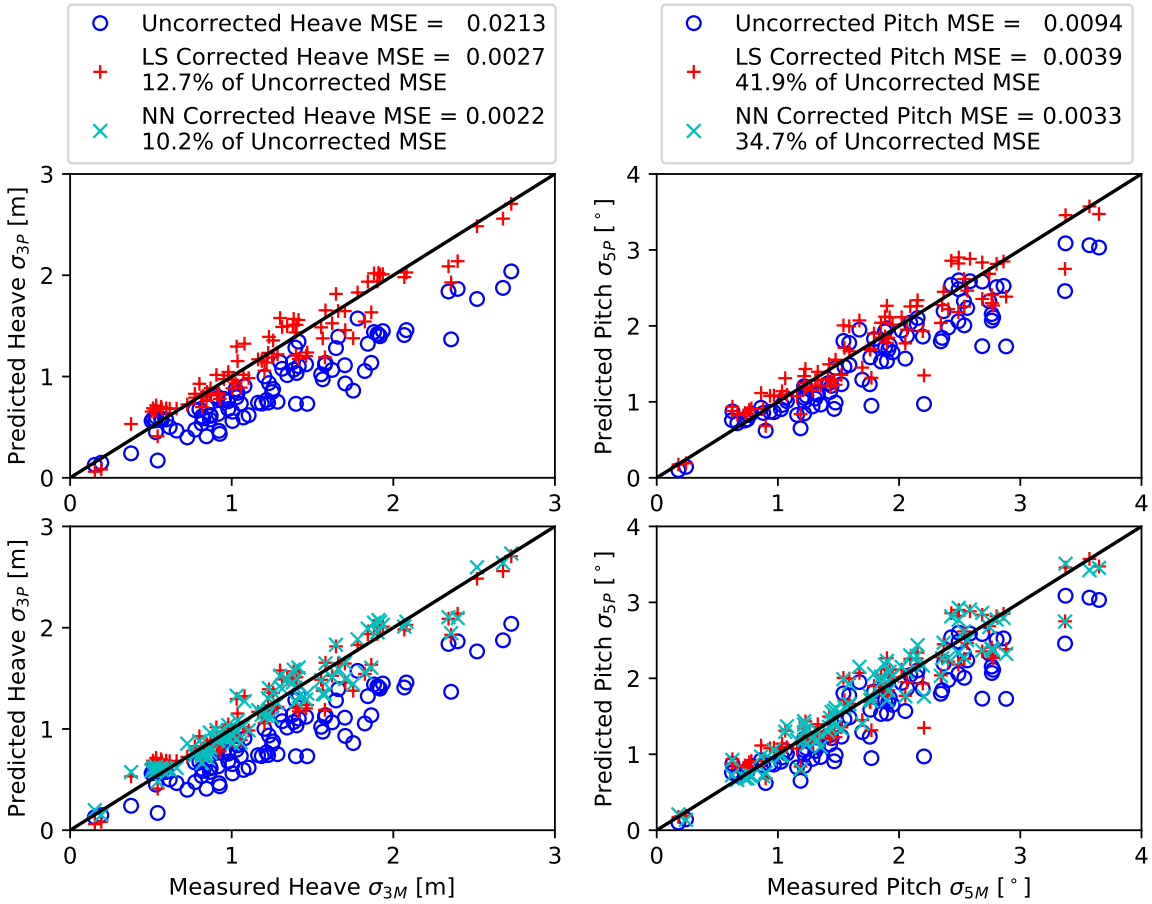

Figure 4.7: Predicted versus measured heave and pitch standard deviations for the randomly partitioned uncorrected, LS corrected, and NN corrected test datasets. The black lines with slope 1.0 represent perfect agreement between predicted and measured values.

chronological partitioning. As mentioned in Section 4.1.3, the diversity of conditions experienced during this cruise was limited, and the addition of data from other cruises may improve the LS and NN correction effectiveness for chronologically partitioned trials. The increased MSE reduction for the 10 trials that used random partitioning likely resulted from increased input variable similarity between the test and train datasets. For example, two successive points with similar combinations of H_s, T_p, U, and θ may have been separated into the train and test datasets through random partitioning. As a result, it is likely that the correction would be more effective for this test data point than it would be for a test data point with a unique combination

of H_s, T_p, θ, and U that was not well-represented in the train dataset. It is expected that an increase in the data available to train these correction models would naturally lead to an improvement in a chronologically split dataset by reducing data sparsity among input variables. These trade-offs between different partitioning methods are revisited in subsequent chapters of this dissertation.

It should be noted that while the differences in MSE reduction between the Best and Worst values of Table 4.2 were relatively small for all of the trials performed, the NN architectures used to yield these results, given in Table 4.1, were not consistent between trials. The diversity of NN architectures that yielded similar MSE reduction, even for the chronologically split data where the test and train datasets did not change between trials, implies that the data used were not particularly sensitive to the NN structure. While NNs are useful tools for linear and nonlinear corrections, their complexity makes it difficult to identify the dependencies between input variables and the sources of uncertainty that cause discrepancies between predicted and measured values. As a result, small changes in the full-scale data collection process or vessel loading and operating conditions could invalidate the trained NN. Application of the LS and NN correction models trained in this work to data from a different cruise of the *R/V Knorr* would provide insight regarding the versatility of these models. Alternatively, the subsequent chapters of this dissertation apply similar models to a much larger dataset resulting from numerous cruises of another research vessel.

4.2.2 *USCGC Bertholf* VBM Corrections

The results of these VBM correction models for the 45° and 5° heading filtered data, originally presented in *Hageman et al.* (2019) and *MARIN* (2019), are shown in Figures 4.8 and 4.9, respectively. In each figure, the left plot is the uncorrected test data, the middle plot is the LS corrected test data overlaid on the uncorrected test data, and the right plot is the NN corrected test data overlaid on the uncorrected

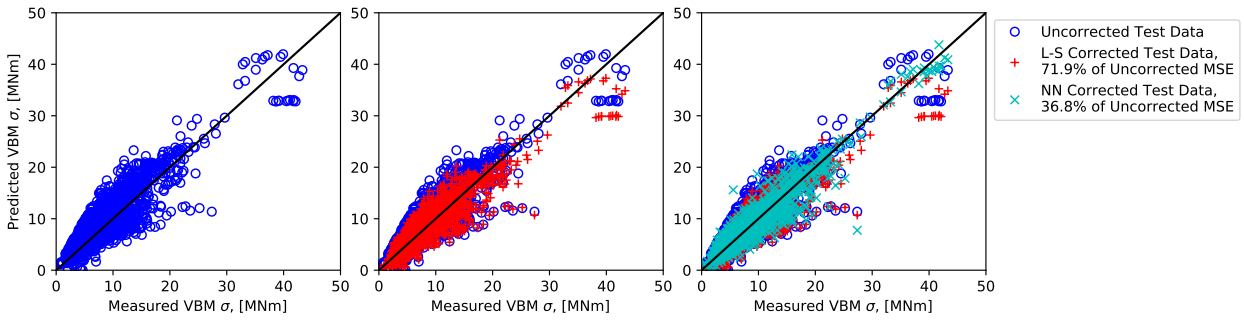

Figure 4.8: Predicted versus measured VBM standard deviation for the uncorrected and corrected 45° heading filtered datasets. The left plot is the uncorrected test data, the middle plot is the LS corrected test data overlaid on the uncorrected test data, and the right plot is the NN corrected test data overlaid on the uncorrected and LS corrected test data. Each test dataset contains 2,025 points.

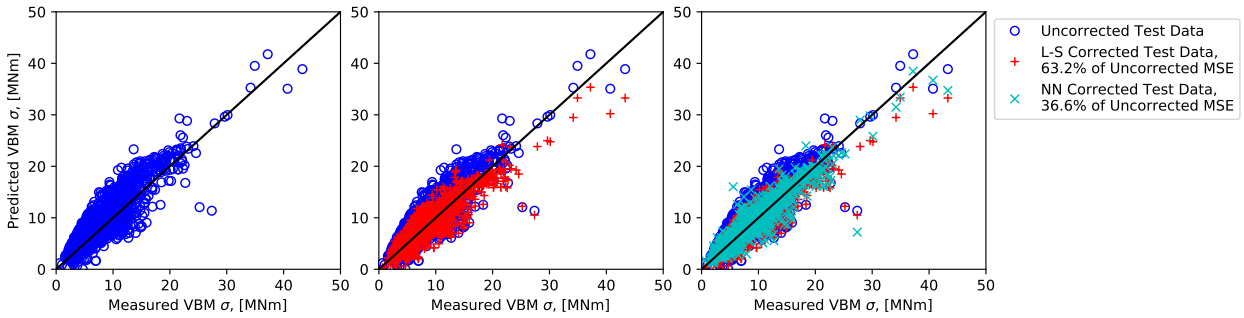

Figure 4.9: Predicted versus measured VBM standard deviation for the uncorrected and corrected 5° heading filtered datasets. The left plot is the uncorrected test data, the middle plot is the LS corrected test data overlaid on the uncorrected test data, and the right plot is the NN corrected test data overlaid on the uncorrected and LS corrected test data. Each test dataset contains 1,371 points.

and LS corrected test data. Note that the NN results were not perfectly consistent between trials of the NN due to underlying randomness in the optimizer used to train the network, and the random partitioning of the dataset also introduced inconsistency between trials for all three datasets presented in Figure 4.8 and Figure 4.9. As mentioned previously, although less formal than the heave and pitch correction study, multiple trials were performed, and the results presented here are representative of the average results of those trials. Recall that the final NN selected (for both the 5° and 45° heading filtered datasets) had five HLs with 60 neurons each. As shown in the legend of Figure 4.8, for the 45° heading filtered data, the LS correction yielded

71.9% of the uncorrected test data MSE, and the NN correction yielded 36.8% of the uncorrected test data MSE. For the 5° heading filtered data, the LS correction yielded 63.2% of the uncorrected test data MSE, and the NN correction yielded 36.6% of the uncorrected test data MSE. The reduction of MSE due to the LS correction and the more significant reduction due to the NN correction are visually evident in both filtered datasets. Comparing Figure 4.8 and Figure 4.9, the 5° heading filter removes many of the points with high error magnitudes present in the 45° heading filtered dataset, but the relative effectiveness of the LS and NN correction methods did not change significantly between filtered datasets.

4.3 Preliminary Correction Conclusions

In the *R/V Knorr* heave and pitch correction study, originally discussed in *Schirmann et al.* (2019a), for both the chronological and random partitioning methods, the LS and NN correction methods successfully reduced the MSE of test datasets over 20 trials. While the NN outperformed the LS correction in all 20 trials, the relative differences between the uncorrected and LS corrected heave and pitch test data were much greater than the differences between the LS and NN corrected test data. These results suggest that the majority of the NN corrections applied here accounted for linear uncertainty in the data. While the data available for these analyses were limited, corrections using chronologically partitioned data did lead to significant reduction in MSE relative to the uncorrected values. Furthermore, the more significant MSE reduction of the randomly partitioned data trials suggests that expanding the dataset used to train these correction models may improve the results of the chronologically partitioned data trials, which more closely model a real-world application of data-driven models for response prediction.

Further demonstrating the power of these data-driven correction approaches, the *USCGC Bertholf* VBM correction study, originally discussed in *Hageman et al.* (2019)

and *MARIN* (2019), showed that the LS and NN correction methods effectively reduced the MSE between measured and predicted VBM standard deviations. For both the 45° and 5° heading filtered dataset, the trained NN outperformed the LS approach and reduced the MSE of a test dataset to under 40% of its original value. Although not provided to the Marine Structures Design Laboratory (MSDL), a similar VBM dataset from the *USCGC Stratton*, a sister of the *Bertholf*, was available to another VALID II JIP participant. The MSDL provided this participant with the two (45° and 5°) *Bertholf*-trained NNs, and their performance for VBM correction was evaluated on the *USCGC Stratton* dataset. Although the results shared with the MSDL were limited, both trained NNs successfully reduced the MSE of the *Stratton* dataset. Note, however, that the reduction in MSE was less significant (∼10% MSE reduction) than it was for the *Bertholf* results discussed here (∼60% MSE reduction). Nevertheless, given that the *Bertholf* test results are not a true, independent performance evaluation for the aforementioned reasons, any reduction in MSE for the *Stratton* dataset shows promise for application of data-driven models across ships in a fleet. As highlighted in Chapters I and II, this implementation of structural health monitoring could aid vessel owners and operators in underway, maintenance, and deployment decisions.

The success of the LS and NN approaches for *R/V Knorr* motion and *USCGC Bertholf* VBM corrections merited further investigations using more intensive data-driven models based on multidirectional wave data. The insights gained from these preliminary experiments, conducted using datasets with significant limitations, formed a solid foundation for rigorous application and evaluation of additional linear and NN approaches with much larger, more accessible datasets in the following chapters.

CHAPTER V

Improved Research Vessel Motion Prediction Models

The LS and NN approaches of the Chapter IV were framed as "correction" models because, in addition to the unidirectional wave data parameters, the model inputs included the uncorrected heave and pitch or VBM amplitude predictions of the initial twin framework. From this correction point of view, and given the computational efficiency of the initial twin's linear, frequency-domain approach, these initial twin framework predictions, or Physics-Based Model Predictions (PBMPs), were logical input variables to the LS and NN models. Furthermore, for the *R/V Knorr* heave and pitch dataset in particular, inclusion of these PBMPs was deemed necessary given the very limited number of training samples available to characterize such a large wave data input space. Given a much larger dataset with a wider array of operating conditions and more detailed, multidirectional wave data, would these PBMPs add any value to the models? The study discussed in this chapter addressed this question and improved upon the preliminary models of Chapter IV using multidirectional wave data and over 16,000 30-minute windows from two operational sister vessels. The following study was originally presented in *Schirmann et al.* (2021), with preliminary discussion in *Schirmann et al.* (2020b). Note that *Schirmann et al.* (2020b) only considered data from one of the two vessels, a total of approximately 13,400 30-minute

windows.

One common criticism of data-driven models, which the work in this chapter sought to combat, is that many implementations abandon the well-understood physics on which traditional methods are based; the models learn from experiences included in the training dataset, but exclude physics-based information that may be useful when a model is applied to novel data. Furthermore, many ML approaches, especially NNs, are often criticized as black boxes whose inner workings are difficult to interpret. These approaches are designed to reduce uncertainty and error between data-driven model predictions and real-world behavior, but the origins of these corrected errors are often untraceable through the model. For example, it can be difficult or impossible to quantify the uncertainty associated with input variables versus the uncertainty due to model form errors, or inadequacies in the underlying prediction model, as defined by *Subramanian and Mahadevan* (2019). As such, the combination of physics-based information and machine learning has garnered significant attention across engineering disciplines in recent years (e.g., *Willcox* (2019), *Raissi et al.* (2019), *Kapusuzoglu and Mahadevan* (2020)).

The potential deficiencies of data-driven models mentioned above can be combated through incorporation of physics-based information. Mentioned in Chapter IV, *Weymouth and Yue* (2013) discussed a framework for physics-based learning models and demonstrated their application to several hydrodynamic modeling challenges. These challenges included extrapolation of sparse experimental RAO data in head seas to Froude numbers unseen in the training data. They also demonstrated that coupling numerical models with measured realizations of a system's state presents an opportunity to use more traditional, computationally efficient numerical models rather than time intensive processes (e.g., potential flow codes versus CFD simulations) to yield similar accuracy. Similarly, a small number of high-fidelity realizations could be used to train these physics-based learning models rather than a large num-

ber of computationally expensive runs, which was demonstrated in the context of bow wave behavior prediction. *Weymouth* (2019) further described this method and demonstrated incorporation of a physical basis into Ridge Regression (RR) models for prediction of roll damping coefficients.

To examine the benefits of including physics-based information in data-driven models, linear ridge regression (RR) and (nonlinear) NN models were constructed, trained, and tested both with and without PBMPs included as input variables. The RR approach employed here was very similar to the previous LS approach, with differences detailed in Section 5.1.8. Similar to the models detailed in Chapter IV, these data-driven predictive models were based on the vessel's speed and time-and-location specific wave model data, with wave directions defined relative to the vessel's heading. Along with the use of multidirectional wave data, another key difference between this study and the preliminary heave and pitch prediction approaches of Chapter IV was the addition of roll motion as a third output variable.

The performances of the RR and NN models with and without PBMPs were compared using two test datasets. One of these test datasets was taken from the same vessel used in model training (with preliminary analyses presented in *Schirmann et al.* (2020b)), and the other was taken from a sister vessel. The application of these trained models to a sister ship was performed to evaluate the versatility of these data-driven models, similar to the experiment mentioned in Chapter IV for VBM correction between the *USCGC Bertholf* and *USCGC Stratton*. More explicitly, if data-driven models perform well between similar ships, they may prove useful for fleet management, and multiple ships could be used to accelerate data collection for model training. Additionally, operating profiles may vary between the two ships, further testing the adaptability of the trained models to unique conditions.

As mentioned in Chapter I, reliable data-driven models could be used by vessel owners and operators to support operational guidance and deployment decisions

through motion forecasting. While most of the data-driven model training and testing discussed here was performed using wave hindcast data due to its continued availability, a small test dataset of wave forecast data was also stored for additional demonstration of the most-successful models. This final experiment demonstrated real-world application of these models for vessel motion forecasting at different time horizons.

Compared to the other vessel motion prediction NN studies mentioned at the beginning of Chapter IV, the novelty of this work lies in its large scope (two sister vessels and over 16,000 real-world, 30-minute time windows), the use of time-and-place specific multidirectional wave model data, and the direct incorporation of physics-based information and investigation of its benefits.

To summarize, this study's purpose was to evaluate ship motion forecasting using data-driven models through realization of the following goals:

1. Demonstrate the use of linear ridge regression (RR) and nonlinear neural network (NN) approaches for vessel motion prediction using hindcast multidirectional wave model data

2. Investigate the potential benefits of incorporating physics-based model predictions (PBMPs) as input to these data-driven models

3. Evaluate model robustness through comparison of predicted responses using forecast and hindcast wave model data

4. Test the versatility of the data-driven models through application to a sister ship of the original vessel

Data from approximately 13,500 30-minute windows, resulting from 50 cruises of the *R/V Neil Armstrong*, were considered for training and testing these ML models for motion prediction, which addressed goals 1,2, and 3. Additionally, another test

dataset generated from over 2,500 30-minute windows, resulting from 18 different cruises of the *R/V Sally Ride*, were used to address the final goal and further support goals 1 and 2.

5.1 Methodology

This section, which follows *Schirmann et al.* (2020b) and *Schirmann et al.* (2021), details the different datasets considered from the sister vessels *R/V Neil Armstrong* and *R/V Sally Ride*, as well as the construction, training, and testing of the linear RR and nonlinear NN models models with and without incorporation of PBMPs as input. Section 5.1.1 gives an overview of the two ships, and Section 5.1.2 details the data collected aboard each vessel and the data processing performed. Sections 5.1.3 and 5.1.4 discuss the the different datasets considered from the *Neil Armstrong* and *Sally Ride*, respectively. Section 5.1.5 describes the multidirectional wave data and their use as input variables to the data-driven models, and Section 5.1.6 outlines use of the initial twin framework to generate PBMPs with these wave data. Section 5.1.7 discusses the data-driven models' output variables and data normalization, and Sections 5.1.8, 5.1.9, and 5.1.10 detail model construction and training. Finally, Section 5.1.11 briefly compares the different datasets from the two vessels.

Figure 5.1 gives an overview of the data processing approach described in this section. Time series data measured aboard the ship included location, time, heading, speed, and motions (heave, pitch, and roll). Data used in model training and testing were processed in 30-minute windows, which were selected such that each window had approximately stationary speed and heading as detailed in Section 5.1.2.3. As shown in Figure 5.1, statistics from each 30-minute window were then used to select time-and-place specific wave data and generate PBMPs. The data types used as model input and the measured motion statistics used in model training and evaluation are also indicated.

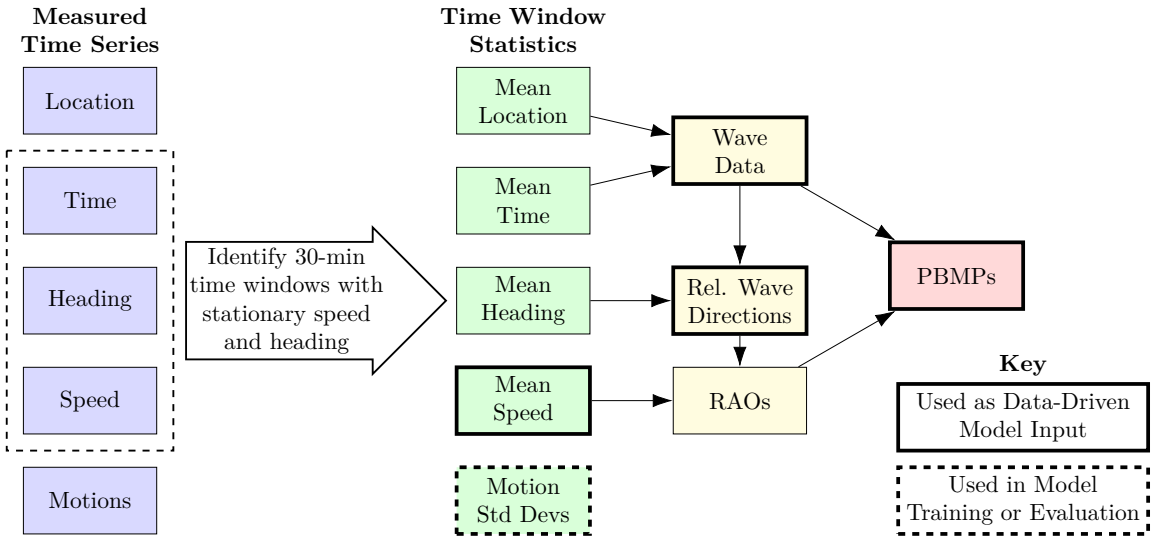

Figure 5.1: An overview of the data processing approach, as well as the data types used as model input variables and for model training and evaluation.

5.1.1 Sister Vessels' Overview

The *R/V Neil Armstrong* (AGOR-27) was the primary vessel for this study, which is operated by Woods Hole Oceanographic Institution (WHOI) in Woods Hole, Falmouth, Massachusetts. Vessel motions data from numerous cruises of the *Neil Armstrong* and corresponding wave model data were separated into two datasets to train and test the prediction models, respectively. All data from the *Neil Armstrong*'s sister vessel, the *R/V Sally Ride* (AGOR-28), were used for additional testing of these models. The *Sally Ride* is operated by the Scripps Institution of Oceanography (SIO) at the University of California, San Diego in San Diego, California. Both ships are owned by the Office of Naval Research (ONR) and were constructed at Dakota Creek Industries in Anacortes, Washington (*U.S. Navy Office of Information*, 2018). The *Neil Armstrong* was delivered in September, 2015 and the *Sally Ride* was delivered in July, 2016. The two ships are pictured in Figure 5.2. The vessels' design speed is 12 knots, and select design characteristics of the vessels are given in Table 5.1. The values in Table 5.1 are based on the hull lines plan provided by WHOI, which was created by the vessels' designers at Guido Perla & Associates. This hull lines plan

Figure 5.2: Unmodified photo of WHOI's *R/V Neil Armstrong* (left) from Ken Kostel, WHOI (`https://commons.wikimedia.org/wiki/File:Armstong_fromthefantail_406493.jpg`) and U.S. Navy photo of the *R/V Sally Ride* (right) from John F. Williams (`https://commons.wikimedia.org/wiki/File:R-V_Sally_Ride_is_currently_underway_conducting_a_series_of_science_verification_cruises._(31534416962).jpg`).

Table 5.1: Design characteristics of the sister ships *R/V Neil Armstrong* and *R/V Sally Ride*. All values are molded and based on the hull lines plan created by Guido Perla & Associates and provided by WHOI.

Design Characteristics	
Length Overall	72.54 m
Waterline Length	70.10 m
Beam	15.24 m
Depth	6.71 m
Draft	4.57 m
Displacement	3043.46 LT
LCG (Aft of FP)	35.38 m
Block Coefficient	0.624

was used for to generate RAOs for the initial twin framework, as detailed in Section 5.1.6. Note that while a theoretical longitudinal center of gravity (LCG) was known from the design specifications, a vertical center of gravity (VCG) was assumed at a value that yielded acceptable transverse stability. The specifics of this VCG selection are revisited in the following discussion. Also consider that vessel loading conditions undoubtedly varied between ships and as a function of time, which would affect both the LCG and VCG; however, these values were not tracked for consideration in this study.

The 50 cruises of the *Neil Armstrong* used to train and test the data-driven models occurred during four calendar years from 2016 through 2019. The additional 18 cruises of the *Sally Ride* used for further model testing occurred during the years 2017 through 2019. Note that these were not the vessels' only voyages during these years, but rather the cruises with available motions data that had appropriate and usable data for model training and testing. As such, some of the considered cruises yielded a small number of usable time windows relative to their duration. The specifications necessary for motions data to be considered appropriate are discussed in Section 5.1.2.3.

5.1.2 Onboard Measurements

Both the *Neil Armstrong* and *Sally Ride* were outfitted with motion measurement instruments that are detailed in the following subsections. The output heave ($j = 3$), pitch ($j = 5$), and roll ($j = 4$) time series recorded by these instruments were processed in 30-minute windows as detailed in Section 5.1.2.3. These motions time series were also accompanied by status flags, or numerical values to indicate normal operation of the measurement instruments. As in Chapter IV, the standard deviations in each motion degree of freedom (DOF) over the 30-minute windows were the representative motion amplitudes that the data-driven models were designed to predict. The three measured target variables are summarized in Table 5.2.

Table 5.2: Measured target variables for data-driven model training.

Index	Variable	Units	Description
0	σ_{3M}	m	Measured Heave Std. Dev.
1	σ_{5M}	deg	Measured Pitch Std. Dev.
2	σ_{4M}	deg	Measured Roll Std. Dev.

5.1.2.1 R/V Neil Armstrong (AR) Measurements

The data from each cruise of the *Neil Armstrong* were accessed through the WHOI Data Library and Archives at `http://dlacruisedata.whoi.edu/AR/`, where `AR` corresponds to the *Armstrong*. Each cruise was assigned a unique cruise identification tag, or cruise ID (e.g., AR29), which corresponds to a subfolder name within the *Armstrong*'s folder. Additional information about these cruises can be found at `https://www.whoi.edu/what-we-do/explore/cruise-planning/cruise-planning-before-the-cruise/cruise-planning-cruise-synopsis/`.

The heave, pitch, and roll motions of the vessel were recorded at approximately 1 Hz for all 50 cruises using the inertial measurement unit (IMU) of an *Applanix POS MV* system that was located in the transducer room and on the centerline of the vessel. Unlike the heave measurements of Chapter IV, which were also influenced by pitch and roll motion, these vertical displacements were assumed to be pure heave based on available calibration information. General information about the vessel's equipment and specifications, including a diagram with the *POS MV*'s location, can be accessed at `https://www.whoi.edu/what-we-do/explore/ships/ships-neil-armstrong/`. The speed, heading, and location (latitude and longitude) of the vessel were also recorded, processed, and provided by WHOI at one-minute intervals.

5.1.2.2 R/V Sally Ride (SR) Measurements

Data from the *Sally Ride*'s cruises were accessed through Rolling Deck to Repository (R2R) at `https://www.rvdata.us/search/vessel/Sally%20Ride`. As with the *Neil Armstrong*, each cruise was assigned a cruise ID (e.g., SR1906), which identified each cruise's datasets in R2R. R2R also provides supporting information for each cruise, such as the start/end dates and ports.

The motion reference unit (MRU) of a Kongsberg Seapath 330+ was used to record heave, pitch, and roll measurements at 5 Hz for all *Sally Ride* cruises considered

in this work. General information and layouts for the *Sally Ride* can be accessed at `https://scripps.ucsd.edu/ships/sally-ride`. Similar to the *Neil Armstrong*, speed, heading, and location data were processed and provided via R2R at one-minute intervals.

5.1.2.3 Measured Data Processing

One goal of this work, originally discussed in *Schirmann et al.* (2020b) and *Schirmann et al.* (2021), was to improve upon the data processing approach of the *R/V Knorr* study of Chapter IV, originally presented in *Schirmann et al.* (2019a), thus increasing the number of time-windows usable for data-driven model training. As mentioned in Chapter IV, for the analyses and frequency-domain motion prediction approach of the initial twin framework, it was necessary to assume that the vessel's operational environment and motions were approximately statistically stationary processes during any usable time window. Recall that in the *R/V Knorr* study of Chapter IV, vessel motions statistics were calculated for three-hour periods that were centered on wave hindcast output times. After calculating motion statistics for each three-hour window, maximum allowable thresholds on the speed and heading standard deviations of the vessel during the three-hour windows were applied. While this was deemed a reasonably effective approach, the relatively large time window length and fixed window centroids resulted in many potentially valuable data being discarded. Furthermore, longer time windows are more vulnerable to changes in wave conditions due to variation in time or the ship's location, detracting from the stationary assumption's validity.

The work discussed in this chapter reduced the window size to 30-minute intervals and modified the measured data windowing approach to ensure that the assumption of statistical stationarity was followed in good faith while increasing the number of available windows for data-driven model training and testing. The new windowing

approach searched for 30-minute time periods in which the speed and heading of the vessel were approximately constant. Beginning with the first 30-minute window for a given cruise, the mean latitude, longitude, speed, and heading of the vessel were calculated. In addition to the requirement that motion data recording was uninterrupted and devoid of irregular status flags, to accept a 30-minute window for use, three requirements needed to be met: the mean latitude and longitude were a minimum of 10 NM from the nearest coast (as determined using the `Matplotlib Basemap Toolkit` in `Python`), the maximum absolute speed deviation from the mean within the 30-minute period was less than 1 knot, and the maximum absolute heading deviation from the mean was less than 15°. These speed and heading deviations were calculated using the preprocessed data provided by WHOI and SIO at one-minute intervals, so maximum deviation thresholds were more appropriate than the standard deviation thresholds employed in the work of Chapter IV. If these three requirements were not met for a given 30-minute period, the 30-minute window was shifted one minute in the positive time direction and the three criteria were checked again. When a 30-minute period did meet all three criteria, and was therefore accepted, the search resumed for new windows beginning with the first minute following the accepted time window.

For all accepted 30-minute windows, the standard deviation of the corresponding heave, pitch, and roll time series were calculated, which match the desired target variables from Table 5.2. Therefore, in each DOF, this windowed time series was comprised of approximately 1,800, 1 Hz measurements for *Neil Armstrong* data, and 9,000, 5 Hz measurements for *Sally Ride* data. Additionally, the mean speed (U) for each time window was stored and provided as an input variable to the data-driven models, as discussed in Section 5.1.5. The vessel's mean heading from each time window was also stored for calculation of wave directions relative to the ship heading.

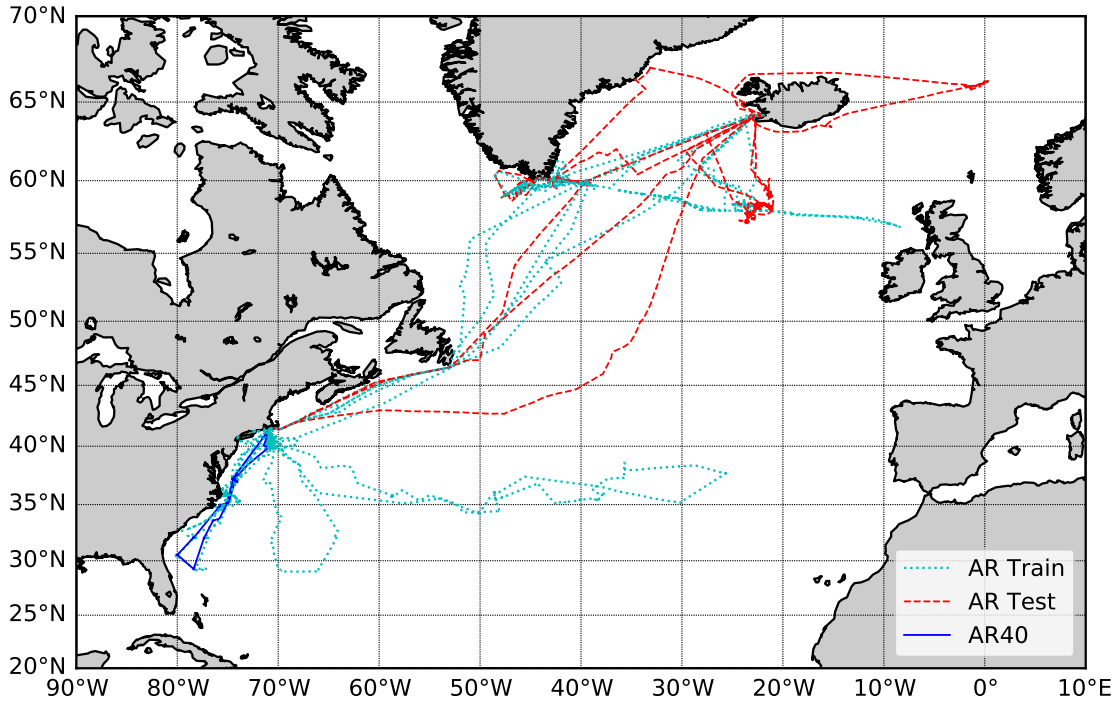

Figure 5.3: Map of the Atlantic Ocean showing the *Neil Armstrong*'s 42 train and 8 test (with AR40 indicated) cruises. The exact train and test data locations used and omissions of cruise portions due to the windowing approach are not reflected in this map.

5.1.3 R/V Neil Armstrong (AR) Train and Test Datasets

Figure 5.3 shows the 50 *Neil Armstrong* cruises in the Atlantic Ocean. Of the 50 *Neil Armstrong* cruises considered, the train dataset, AR Train, was constructed from the first 42, and the test dataset, AR Test, was constructed from 7 subsequent cruises. The final cruise, AR40, was used as a separate test dataset for comparison of data-driven model performances using wave hindcast and forecast data. This comparison is outlined in Section 5.1.5. The map of Figure 5.3 identifies the dataset that each plotted route belongs to. Although the map reflects all location data available from each cruise, the amount of usable time windows from a given cruise often did not span the full route due to the requirements outlined in Section 5.1.2.3.

The AR Test and AR Train datasets are summarized in Table 5.3. In addition

Table 5.3: Summary of the datasets considered in this study. Negative longitude values indicate those west of 0° (e.g., -50° is equivalent to 50°W in Figure 5.3).

Dataset	Lat_N	Lat_S	Lon_E	Lon_W	H_{s50} [m]	H_{s100} [m]	T_{m50} [s]	T_{m100} [s]	Windows
AR Train	64.1	29.0	-8.4	-80.0	1.67	6.41	7.44	13.07	10,008
AR Test	66.9	41.4	0.3	-70.3	1.86	5.84	7.86	12.02	3,384
SR Test	50.3	-6.4	-102.4	94.1	1.43	3.87	9.94	14.94	2,592
AR40	41.3	29.2	-70.9	-80.0					132
120−96h					1.02	2.35	6.73	9.91	
96−72h					1.07	2.34	6.55	10.11	
72−48h					1.07	2.13	6.73	10.61	
48−24h					1.17	2.15	6.87	11.12	
24− 0h					1.17	2.15	6.93	11.12	
Hindcast					1.16	2.33	6.85	11.57	

to the number of windows in each dataset, the median and maximum significant wave heights, H_{s50} and H_{s100}, and the median and maximum mean wave periods, T_{m50} and T_{m100}, of each dataset are included for reference. The source of these wave parameters are discussed in Section 5.1.5. Table 5.3 also includes the northernmost and southernmost latitudes, Lat_N and Lat_S, and the easternmost and westernmost longitudes, Lon_E and Lon_W, of the usable windows in each dataset. These details are included for the 42 cruises of AR Train and the 7 cruises of AR Test individually in Tables 5.4 and 5.5, respectively. Tables 5.4 and 5.5 also include the month that each cruise began.

5.1.4 R/V Sally Ride (SR) Test Dataset

The 18 cruises of the *Sally Ride* used for further model testing are mapped in Figure 5.4. As in Figure 5.3, Figure 5.4 reflects all location data from each cruise rather than the exact time windows with usable data. For instance, less than half of the transpacific voyage shown in Figure 5.4 yielded motions data with acceptable status flags, as mentioned in Section 5.1.2.2. The SR Test dataset is outlined in Table 5.3, with the 18 individual cruises detailed in Table 5.6.

Table 5.4: Summary of the 42 *Neil Armstrong* cruises considered in the AR Train dataset, which are mapped in Figure 5.3. Negative longitude values indicate those west of 0° (e.g., -50° is equivalent to 50°W in Figure 5.3).

Start Month	Cruise ID	Lat_N	Lat_S	Lon_E	Lon_W	H_{s50} [m]	H_{s100} [m]	T_{m50} [s]	T_{m100} [s]	Windows
05/2016	AR3	41.3	39.8	-70.5	-71.3	1.78	2.91	6.65	8.40	197
	AR4-A	41.3	39.9	-70.8	-70.9	0.97	1.50	7.35	9.79	78
	AR4-B	41.3	39.8	-70.3	-70.9	1.04	2.70	7.26	8.19	152
	AR4-C	41.3	39.8	-70.5	-71.0	1.13	1.41	6.02	6.98	121
06/2016	AR5	41.3	39.2	-69.5	-70.9	1.47	2.35	6.04	6.65	96
	AR6	41.3	39.8	-70.6	-70.9	1.34	1.72	7.53	8.12	17
08/2016	AR7-02	64.1	47.7	-23.2	-52.5	1.63	3.33	7.36	10.22	820
09/2016	AR7-03	46.4	41.4	-53.3	-70.3	1.34	2.15	6.44	8.11	149
	AR8A	41.3	39.8	-70.6	-71.1	1.58	3.80	7.14	8.24	164
10/2016	AR8B	41.3	39.8	-70.4	-71.1	1.65	3.10	8.83	9.74	151
11/2016	AR9-01	41.2	32.7	-71.0	-79.5	0.96	2.68	6.88	8.85	115
02/2017	AR9-03	41.3	32.9	-70.9	-78.8	1.11	1.83	6.05	7.56	123
03/2017	AR10	41.3	40.2	-70.4	-71.3	2.18	3.24	6.32	7.37	92
	AR11	41.3	40.4	-70.5	-71.3	2.03	3.27	6.01	7.65	46
	AR12	41.3	40.4	-70.4	-71.4	1.46	3.14	6.38	8.69	88
	AR13	41.3	40.3	-70.4	-71.2	1.62	2.34	6.82	9.73	82
04/2017	AR14	41.3	40.5	-70.5	-71.0	1.70	3.35	7.08	7.75	50
	AR15	41.3	34.6	-70.7	-76.0	1.80	4.43	6.70	9.29	326
05/2017	AR16	41.3	29.0	-64.1	-71.5	1.64	4.68	7.82	9.37	584
	AR17-01	41.3	40.1	-71.1	-73.4	0.92	1.00	6.45	6.81	24
	AR17-02	41.4	40.4	-70.8	-73.7	1.59	1.81	7.59	7.73	27
	AR18A	41.3	39.9	-70.8	-70.9	1.27	1.62	8.28	8.94	85
06/2017	AR18B	41.2	39.8	-70.6	-71.1	1.58	2.55	6.82	8.28	126
	AR18C	41.3	40.1	-70.6	-70.9	1.03	2.18	7.18	8.50	72
	AR19	41.3	41.2	-71.0	-71.3	1.68	1.85	5.46	5.59	10
07/2017	AR20	41.3	39.6	-70.7	-71.4	1.24	1.74	6.73	8.32	146
	AR21	60.1	41.4	-39.1	-70.3	1.37	3.01	7.37	10.15	1047
09/2017	AR22	41.3	40.0	-70.8	-71.0	1.23	1.54	5.07	7.22	19
	AR23-01	41.2	34.2	-26.0	-70.9	2.04	3.22	10.02	11.82	736
10/2017	AR23-02	41.1	34.3	-25.8	-70.8	1.66	3.84	8.24	12.85	406
	AR24-A	41.2	39.9	-70.8	-70.9	2.24	3.97	7.05	8.93	57
	AR24-B	41.4	39.8	-70.6	-70.9	1.94	4.26	7.82	11.57	143
11/2017	AR24-C	41.3	39.9	-70.8	-71.0	1.93	3.32	7.32	10.07	153
	AR25	41.4	29.1	-70.8	-80.0	1.50	3.05	7.14	9.54	533
01/2018	AR26	41.3	34.5	-70.9	-76.1	1.73	3.96	6.81	9.51	299
03/2018	AR28-A	41.4	39.9	-70.8	-70.9	1.46	4.29	6.87	12.88	219
04/2018	AR28-B	41.4	39.8	-70.8	-70.9	1.53	2.99	6.87	8.79	146
	AR29	41.3	39.6	-70.3	-71.1	1.63	4.60	7.12	9.96	110
05/2018	AR30-01	64.0	41.4	-23.1	-70.2	2.62	5.37	8.12	10.04	388
	AR30-02	64.0	58.0	-22.2	-28.0	3.31	6.41	9.02	13.07	389
06/2018	AR30-03	64.1	59.7	-23.1	-41.3	2.00	4.19	7.75	10.38	603
07/2018	AR30-04	64.0	56.9	-8.4	-36.5	1.66	3.24	7.08	9.98	819
		64.1	29.0	-8.4	-80.0	1.67	6.41	7.44	13.07	**10,008**

Table 5.5: Summary of the 7 *Neil Armstrong* cruises considered in the AR Test dataset, which are mapped in Figure 5.3. Negative longitude values indicate those west of 0° (e.g., -50° is equivalent to 50°W in Figure 5.3).

Start Month	Cruise ID	Lat_N	Lat_S	Lon_E	Lon_W	H_{s50} [m]	H_{s100} [m]	T_{m50} [s]	T_{m100} [s]	Windows
08/2018	AR30-05	66.8	63.1	0.3	-25.0	1.38	2.46	7.00	9.76	431
	AR30-06	66.9	58.6	-22.5	-48.5	2.19	5.83	7.68	12.02	620
10/2018	AR30-07	64.0	41.4	-23.1	-70.3	3.04	5.84	8.85	10.90	365
04/2019	AR35-01	64.0	41.4	-23.1	-70.2	2.69	5.19	7.97	9.30	467
05/2019	AR35-02	63.9	58.9	-21.0	-23.2	1.28	4.14	8.13	10.84	259
	AR35-03	64.0	57.1	-21.0	-29.5	2.00	3.89	7.83	10.63	535
08/2019	AR35-05	64.1	41.4	-23.1	-70.3	1.36	2.89	7.99	10.38	707
		66.9	41.4	0.3	-70.3	1.86	5.84	7.86	12.02	**3,384**

Table 5.6: Summary of the 18 *Sally Ride* cruises considered in the SR Test dataset, which are mapped in Figure 5.4. Negative longitude values indicate those west of 0°, but still east of 180°. For example, -170° is equivalent to 170°W, and 170° is equivalent to 170°E in Figure 5.4).

Start Month	Cruise ID	Lat_N	Lat_S	Lon_E	Lon_W	H_{s50} [m]	H_{s100} [m]	T_{m50} [s]	T_{m100} [s]	Windows
01/2017	SR1701	33.0	32.6	-118.2	-119.1	0.93	1.05	9.99	11.84	32
03/2017	SR1706	32.7	32.7	-118.9	-119.1	1.37	1.45	11.09	11.29	2
03/2017	SR1708	32.9	32.9	-117.5	-117.5	0.99	0.99	11.66	11.66	1
03/2017	SR1709	33.3	32.7	-117.5	-119.0	0.99	2.46	10.49	12.22	29
09/2017	SR1714	35.0	32.7	-117.5	-120.8	1.51	2.62	7.19	10.64	45
10/2017	SR1716	34.9	32.7	-117.5	-120.9	1.04	2.81	10.90	14.94	61
11/2017	SR1717	34.9	29.9	-117.4	-124.3	1.55	2.58	10.63	12.56	263
12/2017	SR1718	34.3	32.7	-117.5	-120.0	0.90	2.00	10.04	12.48	46
03/2018	SR1804	32.7	32.7	-117.5	-117.5	0.99	0.99	10.28	10.28	1
03/2018	SR1805	22.3	10.1	-102.4	-115.3	1.78	2.42	10.73	13.17	202
06/2018	SR1809	44.3	32.7	-117.5	-124.8	1.47	2.37	8.32	12.81	95
07/2018	SR1811	50.3	44.7	-124.5	-145.2	1.38	2.26	8.21	10.91	309
10/2018	SR1814	41.7	33.2	-118.8	-124.7	2.08	2.30	11.09	12.31	55
10/2018	SR1815	35.0	29.9	-117.7	-124.3	1.86	2.90	10.62	14.70	246
03/2019	SR1904	33.8	32.7	-117.6	-120.3	1.49	2.05	8.50	9.31	24
03/2019	SR1906	32.6	22.9	-117.4	-157.4	2.62	3.87	11.95	14.84	308
05/2019	SR1908	13.4	-6.4	144.5	94.1	0.90	1.89	7.56	14.48	684
10/2019	SR1914	24.2	7.3	134.8	122.9	0.99	2.43	8.76	12.98	189
		50.3	-6.4	-102.4	94.1	1.43	3.87	9.94	14.94	**2,592**

5.1.5 CMEMS Wave Data and Input Variables

The wave data source used in this study came from the European Union's Copernicus Marine Environment Monitoring Service (CMEMS), which was included in the

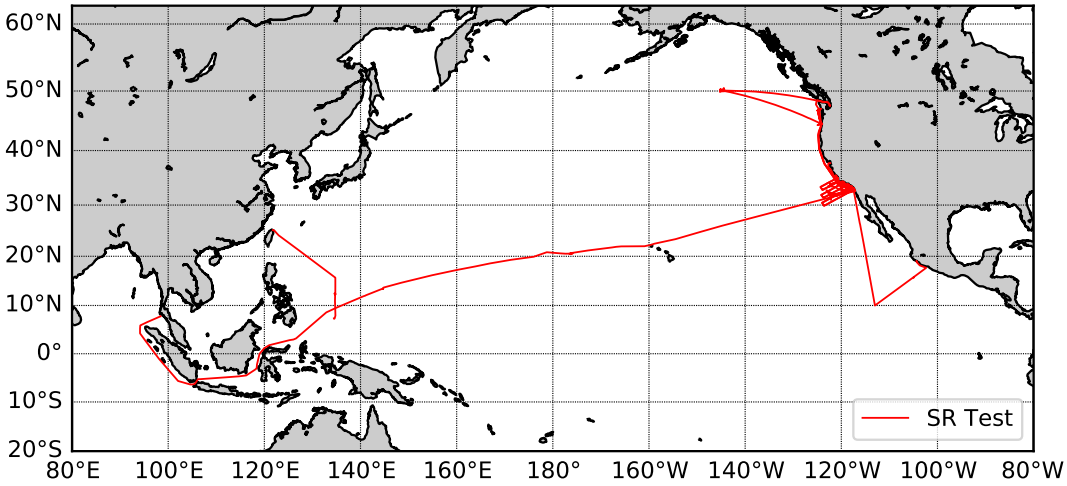

Figure 5.4: Map of the Pacific Ocean showing the *Sally Ride*'s 18 test cruises. The exact train and test data locations used and omissions of cruise portions due to the windowing approach are not reflected in this map. Also note the change in latitude/longitude scales relative to Figure 5.3.

wave data source selection study of Chapter III. Once again, the wave data for all cruises were downloaded through the Global Ocean Waves Analysis and Forecasting Product (GLOBAL_ANALYSIS_FORECAST_WAV_001_027). Recall that for this source, multidirectional wave data were provided at three-hour time intervals for a majority of the world's oceans in latitude and longitude increments of 5 arcminutes $(\frac{1}{12}^{\circ})$. Unlike the work of Chapter IV, due to the relatively high spatial resolution of the wave data used in this study and the selected 30-minute window for the onboard measurements in contrast to the three-hour wave hindcast intervals, all wave data parameters were interpolated in three dimensions: latitude, longitude, and time. The specific interpolation approach employed here is described in the following paragraph.

Table 5.7: Data-driven model input variables. The primary data source(s) used to generate each variable is(are) included. Although not explicitly listed, note that time and location information from WHOI/SIO's onboard measurements were also required to determine all of the CMEMS wave data variables (indices 1-15).

Index	Variable	Units	Description	Source(s)
0	U	m/s	Vessel Speed	WHOI/SIO
1	$VHM0$ (H_s)	m	Significant Wave Height	CMEMS
2	$VTM10$ (T_m)	s	Mean Wave Period (m_0 and m_1)	CMEMS
3	$VTM02$	s	Mean Wave Period (m_0 and m_2)	CMEMS
4	$VTPK$	s	Wave Period at Spectral Peak	CMEMS
5	$VMDR_{Rel}$	deg	Mean Wave Direction (Relative to Ship Heading)	CMEMS/WHOI/SIO
6	$VPED_{Rel}$	deg	Wave Direction at Spectral Peak (Relative to Ship Heading)	CMEMS/WHOI/SIO
7	$VHM0_{WW}$ ($H_{s,WW}$)	m	Wind Waves Significant Wave Height	CMEMS
8	$VTM01_{WW}$ ($T_{m,WW}$)	s	Wind Waves Mean Wave Period (m_0 and m_1)	CMEMS
9	$VTMDR_{Rel,WW}$	deg	Wind Waves Mean Wave Direction (Relative to Ship Heading)	CMEMS/WHOI/SIO
10	$VHM0_{SW1}$ ($H_{s,SW1}$)	m	Primary Swell Significant Wave Height	CMEMS
11	$VTM01_{SW1}$ ($T_{m,SW1}$)	s	Primary Swell Mean Wave Period (m_0 and m_1)	CMEMS
12	$VTMDR_{Rel,SW1}$	deg	Primary Swell Mean Wave Direction (Relative to Ship Heading)	CMEMS/WHOI/SIO
13	$VHM0_{SW2}$ ($H_{s,SW2}$)	m	Secondary Swell Significant Wave Height	CMEMS
14	$VTM01_{SW2}$ ($T_{m,SW2}$)	s	Secondary Swell Mean Wave Period (m_0 and m_1)	CMEMS
15	$VTMDR_{Rel,SW2}$	deg	Secondary Swell Mean Wave Direction (Relative to Ship Heading)	CMEMS/WHOI/SIO
16	$\sigma_{3P,WW}$	m	Predicted Heave Std. Dev. Due to Wind Waves	SHIPMO/CMEMS/WHOI/SIO
17	$\sigma_{5P,WW}$	deg	Predicted Pitch Std. Dev. Due to Wind Waves	SHIPMO/CMEMS/WHOI/SIO
18	$\sigma_{4P,WW}$	deg	Predicted Roll Std. Dev. Due to Wind Waves	SHIPMO/CMEMS/WHOI/SIO
19	$\sigma_{3P,SW1}$	m	Predicted Heave Std. Dev. Due to Primary Swell	SHIPMO/CMEMS/WHOI/SIO
20	$\sigma_{5P,SW1}$	deg	Predicted Pitch Std. Dev. Due to Primary Swell	SHIPMO/CMEMS/WHOI/SIO
21	$\sigma_{4P,SW1}$	deg	Predicted Roll Std. Dev. Due to Primary Swell	SHIPMO/CMEMS/WHOI/SIO
22	$\sigma_{3P,SW2}$	m	Predicted Heave Std. Dev. Due to Secondary Swell	SHIPMO/CMEMS/WHOI/SIO
23	$\sigma_{5P,SW2}$	deg	Predicted Pitch Std. Dev. Due to Secondary Swell	SHIPMO/CMEMS/WHOI/SIO
24	$\sigma_{4P,SW2}$	deg	Predicted Roll Std. Dev. Due to Secondary Swell	SHIPMO/CMEMS/WHOI/SIO
25	σ_{3P}	m	Predicted Heave Std. Dev. Due to All Three Wave Partitions	SHIPMO/CMEMS/WHOI/SIO
26	σ_{5P}	deg	Predicted Pitch Std. Dev. Due to All Three Wave Partitions	SHIPMO/CMEMS/WHOI/SIO
27	σ_{4P}	deg	Predicted Roll Std. Dev. Due to All Three Wave Partitions	SHIPMO/CMEMS/WHOI/SIO

For a given measured data window, interpolated wave data values were calculated at the mean latitude, longitude, and time from a 30-minute measurement period. To do so, eight wave data points, each a corner of the rectangular prism containing these mean value coordinates in 3D (latitude, longitude, and time) space, were used to calculate linearly interpolated wave data parameters. For each time and location, 15 parameter types were extracted from the CMEMS wave hindcast and appropriately interpolated. These 15 parameters included several significant wave height, period, and direction metrics, and are included in Table 5.7 as indices 1-15. The names of these 15 variables correspond to the names assigned by CMEMS as described in *Fernandez and Aouf* (2018). Note that select variables include alternate names in parentheses that match more common notation used in other sections of this discussion. Also note that special care was taken to avoid issues with circularity (e.g., transition from 359° to 0°) when calculating means and interpolating directional parameters.

As shown in Table 5.7 and discussed in Chapter III, in addition to more general parameters for the wave system as a whole, the CMEMS data provided information regarding the multidirectionality of the waves through significant wave height, mean wave period, and mean wave direction parameters for three different wave partitions: wind waves (WW), primary swell ($SW1$), and secondary swell ($SW2$). Note the inclusion of multiple period metrics that are each defined in Table 5.7, where m_0, m_1, and m_2 correspond to the zeroth, first, and second moments of the wave spectrum, respectively. As mentioned in the Section 5.1.2.3, all of the CMEMS direction parameters were adjusted such that they were relative to the vessel's heading, with 180° corresponding to head seas and 0° corresponding to following seas. For this reason, both CMEMS and WHOI/SIO are listed as data sources for these direction parameters in Table 5.7. Due to the port-starboard symmetry of the vessel, the seakeeping model RAOs described in the next section were identical between wave systems com-

ing from the port and starboard sides. Therefore, any relative wave directions on $(180^{\circ}, 360^{\circ})$ were flipped to be on $[0^{\circ}, 180^{\circ}]$ for the purposes of data-driven model training. Note that in applications which require more rigorous handling of phase relationships, such as time-domain analysis, distinction between the port and starboard sides would be necessary.

One drawback of the wave data interpolation approach employed here is that CMEMS' identification of the three wave partitions was not necessarily consistent between the eight corners in latitude, longitude, and time space. For example, the designation of primary swell at each interpolation corner was assigned to the swell partition with the greater significant wave height. If the hierarchy of these swell partition wave heights switched between interpolation corners, the identification of primary and secondary swell would be inconsistent, which would yield unreliable interpolated values, especially with regard to the wave partitions' mean directions. While these instances were likely uncommon, it is important to note that this interpolation method was imperfect, and data-driven model structures that circumvent this issue are detailed in Chapter VI.

In addition to the wave hindcast data used for all work with the AR Train, AR Test, and SR Test datasets, CMEMS provides wave forecast data for up to 120 hours (5 days) in the future, updated once daily (*Fernandez and Aouf*, 2018). Therefore, for any given time of interest, there are five different forecast updates prior. Of these five updates, one update will fall in each of the following time horizon windows, defined as the time between forecast file upload and the time of interest: 120−96h, 96−72h, 72−48h, 48−24h, and 24−0h. As mentioned in Section 5.1.3, the purpose of the final *Neil Armstrong* cruise, AR40, was to compare the performance of the data-driven models using these five forecast wave datasets and the corresponding hindcast wave dataset. Note that CMEMS wave forecast files are only available for a short period surrounding the applicable time windows, after which only the corresponding

hindcast files are available. As such, the storage of these forecast files was planned and executed specifically for this single cruise. A summary of AR40, which yielded 132 windows, is included in Table 5.3. The wave data corresponding to the five forecast time horizon windows and the hindcast data are also summarized in Table 5.3. The results of this comparison study are discussed in Section 5.2.4.

5.1.6 Physics-Based Model Predictions (PBMPs)

As in previous chapters, `SHIPMO.BM` was used to calculate RAOs for the sister vessels, which were then used in the initial twin framework to generate PBMPs. These RAOs were generated for speeds of 0 to 15 knots in increments of 1 knot and relative wave headings of 0 to 355° in increments of 5°. In the same fashion as the initial twin demonstration of Chapter II, Bretschneider spectra were constructed for each of the three wave partitions using the significant wave height and mean period variables identified in Table 5.7, and the resultant heave, pitch, and roll responses were calculated. In contrast to the *R/V Knorr* study of Chapter IV, which used unidirectional wave data, no spreading was applied to these multidirectional wave partitions. As shown by indices 16-27 of Table 5.7, the representative heave, pitch, and roll amplitudes due to each wave partition ($\sigma_{jP,WW}$, $\sigma_{jP,SW1}$, and $\sigma_{jP,SW2}$), as well as the resultant amplitudes due to all three wave partitions (σ_{jP}), were input to the data-driven models as PBMPs when applicable.

As mentioned previously, the VCG of the vessel was unknown. As such, the VCG was assumed at the waterline because it yielded a marginally stable vessel for the purposes of roll prediction. In addition to the typical inaccuracy of empirical models for roll prediction, this uninformed and somewhat arbitrary VCG selection means that the uncorrected roll PBMPs may be incorrect owing to different GM values. Despite the likely low quality of these roll PBMPs, they were still included to test whether the values would aid the data-driven models in any way.

5.1.7 Model Output Variables and Data Normalization

The three output variables of the data-driven models are listed in Table 5.8, which correspond to the measured target variables used to train the models given in Table 5.2. Because machine learning models are often sensitive to significant differences in

Table 5.8: Data-driven model output variables and their normalization factors, which were the maximum values recorded in the AR Train measurements.

Index	Variable	Units	Norm. Factor	Description
0	σ_{3RR} or σ_{3NN}	m	1.822	Predicted Heave Std. Dev.
1	σ_{5RR} or σ_{5NN}	deg	2.963	Predicted Pitch Std. Dev.
2	σ_{4RR} or σ_{4NN}	deg	6.567	Predicted Roll Std. Dev.

magnitudes between variables, all of the input and output variables were normalized to take values on $[0, 1]$, approximately, with deviations from this range explained in the following. For all directional input variables listed in Table 5.7, which were previously adjusted to be on $[0^\circ, 180^\circ]$, this normalization was achieved by dividing all values by 180°. For all significant wave height and period metrics, as well as speed, the maximum value seen in AR Train for a specific variable was used to normalize all of the data for that variable in all datasets. As such, some of the normalized values in the test datasets exceeded a value of 1 if larger magnitudes were present in the test set than the train set; however, using the maximum of all data (train or test), as was done in the preliminary experiments of Chapter IV, would technically compromise the integrity of the test dataset.

For the remaining input variables and all output variables, which were all representative amplitude metrics for heave, pitch, or roll, the maximum measured value in AR Train for each DOF was used to normalize all data in that DOF. For example, the maximum measured heave standard deviation (σ_{3M}) in AR Train, which resulted from all wave partitions, was used to normalize all input PBMPs of heave standard deviations due to wind waves ($\sigma_{3P,WW}$). These heave, pitch, and roll normalization

factors are included in Table 5.8. Note that because the PBMPs could be greater than the measured values, there were many instances in which both train and test normalized values were greater than 1; however, normalization was still important such that the ranges of different variables were more similar than prior to normalization.

5.1.8 Ridge Regression (RR)

The linear ridge regression (RR) models employed in this work were constructed and trained using `Python`'s `scikit-learn` library (*Pedregosa et al.*, 2011). This RR approach is very similar to the linear LS approach employed in Chapter IV. Given N train data samples, the input data are organized into an N-row matrix, $\mathbf{X}_{Train}$, and the corresponding measured data are organized into a separate N-row matrix, $\mathbf{Y}_{Train}$. In this work, $\mathbf{X}_{Train}$ had dimensions $N \times 17$ when PBMPs were excluded and $N \times 29$ when PBMPs were included. In each case, respectively, the first 16 or 28 columns corresponded to the input variables given in Table 5.7, and the one additional column was a vector of ones for multiplication with a bias term. The measured output matrix, $\mathbf{Y}_{Train}$, had dimensions $N \times 3$ with the 3 columns corresponding to the measured heave, pitch, and roll values outlined in Table 5.2. The minimization problem addressed in RR is posed in Equation 5.1, which shows that $\mathbf{X}_{Train}$ is multiplied by a matrix of weights, $\mathbf{W}$. In this work, $\mathbf{W}$ had dimensions 17×3 without PBMPs or 29×3 with PBMPs.

$$\mathbf{W}_{RR} = \arg\min_{\mathbf{W}} ||\mathbf{X}_{Train}\mathbf{W} - \mathbf{Y}_{Train}||_F^2 + \lambda||\mathbf{W}||_F^2 \tag{5.1}$$

In Equation 5.1, $||\cdot||_F^2$ denotes the squared Frobenius norm, which is the sum of all squared matrix elements. The key difference between the linear LS approach employed in Chapter IV and this linear RR approach is the incorporation of an $L2$ weight regularization parameter, λ, to formally mitigate overfitting of the train data

as shown in the second term of Equation 5.1. This regularization approach penalizes large weight values in matrix $\mathbf{W}$. Large weight values are often undesirable because they can cause a model to rely too heavily on individual input variables. Note that the final row of $\mathbf{W}$, which contains three bias terms corresponding to each DOF, is excluded from this penalization. Overfitting can lead to poor generalization, or performance when a model is applied to an unseen test data set. The tuning of λ using train data is discussed in Section 5.1.10. Given λ, the solution to this minimization problem is given in Equation 5.2.

$$\mathbf{W}_{RR} = [\mathbf{X}_{Train}^T \mathbf{X}_{Train} + \lambda \mathbf{I}]^{-1} \mathbf{X}_{Train}^T \mathbf{Y}_{Train} \tag{5.2}$$

Using the trained weight matrix, $\mathbf{W}_{RR}$, and a matrix of test input data, $\mathbf{X}_{Test}$, with dimensions $M \times 17$ without PBMPs or $M \times 29$ with PBMPs, where M is the number of samples in the test set, RR test predictions were calculated as shown in Equation 5.3 and compared to the known measured test values for evaluation of the model.

$$\mathbf{Y}_{RR,Test} = \mathbf{X}_{Test} \mathbf{W}_{RR} \tag{5.3}$$

The RR model's performance on the test dataset was then evaluated as shown in Equation 5.4.

$$MSE_{RR,Test} = \frac{1}{3} ||\mathbf{Y}_{RR,Test} - \mathbf{Y}_{Test}||_F^2 \tag{5.4}$$

Analogous to the metric employed in Chapter IV, this MSE metric is the mean of the heave, pitch, and roll MSE values, which are also analyzed independently in the Section 5.2.

5.1.9 Artificial Neural Network (NN)

NN construction and training was performed using `Python`'s `Keras` library with the `TensorFlow` backend (*Chollet et al.*, 2015). The NNs employed in this study were feed-forward networks composed of an input layer with one neuron corresponding to each of the input variables (16 or 28 with or without PBMPs, respectively), multiple hidden layers with varying numbers of neurons, and an output layer with three neurons corresponding to the heave, pitch, and roll output variables given in Table 5.8. NN training was performed using the MSE loss function, Adam optimizer, and a batch size of 50, which was found to yield similar results to a batch size of 1 with a significant reduction in computation time. The average MSE across heave, pitch, and roll was used as the NN scoring metric during training, analogous to Equation 5.4. Note that several single-output NNs (e.g., roll only) were experimented with in the preliminary stages of this study, which yielded no notable performance improvement over these multi-output NNs when applied to a test dataset.

Because the goal of these networks was regression with input and output variables taking possible values on $[0, \infty)$, both the input and output layer neurons used linear activation functions. The hidden layers (HLs) used neurons with the rectified linear unit (ReLU) activation function, which allowed the NN to model nonlinear relationships between variables unlike the linear RR approach. Similar to the RR approach and in contrast to the NNs of Chapter IV, formal incorporation of $L2$ weight regularization between layers of the NN with parameter λ was implemented to mitigate overfitting, which penalized large weight values via the same mechanism as RR. Weight regularization was not present in the preliminary studies of Chapter IV. The HL architecture (i.e number of HLs and neurons per HL) and λ were determined via cross-validation as described in the following section. For more detailed background information on NNs, please see relevant textbooks such as *Bishop* (2006) and *Russell and Norvig* (2010).

5.1.10 Model Training and Cross-Validation (CV)

For both data-driven approaches, 5-fold cross-validation (CV), described below, was applied to tune model parameters using `GridSearchCV` of `Python`'s `scikit-learn` library (*Pedregosa et al.*, 2011). Due to the relatively high computational cost of NN training, the following cross-validation procedures were performed for NNs using computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor. In this work, the parameters that were formally tuned were λ for both the RR and NN approaches and the HL architecture for the NN only. Due to the previously mentioned computational costs, other NN parameters, such as HL activation functions, batch size, and optimizer, were informally experimented with to arrive at their selected values. For the same reason, note that while significant effort was put toward selecting evenly spread parameters in the search for optimally trained models, the possibilities for these parameters were infinite. Therefore, selectivity was necessary, especially for the HL architectures of the NN.

In 5-fold CV, the train dataset is divided into five subsets of approximately equal size. For a given model and all combinations of candidate parameters specified by the user, training is performed using four of the five subsets and then applied to the one remaining, or held out, subset. The MSE between the model's predictions for the held out subset and the corresponding measured values is stored. This operation is then repeated four more times with the same model parameters such that each of the five subsets is held out once. The average of the five MSE losses is calculated and stored as the average CV loss for a given combination of model parameters. This process is then repeated for all combinations of candidate model parameters specified by the user. After all combinations have been tested, the combination of parameters that yielded the lowest average CV loss is selected as the best model from the user-specified candidates. This best model is then reconstructed and retrained using all of the train data (from all five subsets), at which point model training is complete and

final model evaluation can begin using the unseen test set(s).

Separation of the train dataset into five subsets for 5-fold CV was performed without shuffling the data; i.e., the first subset was the first 20% of the train data samples chronologically, the second subset was the next 20% of train data samples chronologically, etc. This method of partitioning was selected to avoid separating nearly identical samples into multiple CV subsets, especially given the relatively short 30-minute duration employed here. If nearly identical samples collected in sequence were separated into multiple subsets, overfitting may occur despite the regularization parameter, especially if the sequential samples were outliers. Overfitting would cause the model to have poor generalization, yielding inadequate performance when applied to the test data. On the contrary, it is also possible that random shuffling may improve the model's generalization if input data distributions of chronologically partitioned subsets are drastically different. In general, this is a very complex issue that depends heavily on the available dataset for a given project. As such, given the large number of samples available to train the models in this work, the train data was not shuffled prior to partitioning into CV subsets for all results presented here.

Although not presented, some experiments were performed with random shuffling applied before partitioning the train data into CV subsets. As expected, this random partitioning did lead to lower average CV losses due to increased similarity between the five CV subsets; however, because this similarity was somewhat artificial and did not hold between the train and test sets, with shuffling, both the RR and NN models performed similarly or slightly worse than the results presented in the following section where no shuffling was applied. The complexities of CV subset partitioning, specifically in the context of these predictive data-driven models that are trained using (often) sequential time windows, are revisited in Chapter VI.

5.1.11 Comparison of Train and Test Datasets

In an ideal scenario for data-driven models, the train dataset would contain samples that collectively characterize the input data space occupied by test samples. Assuming reliable measurements used to train and test the models, this similarity between train and test datasets should yield very reliable predictions on the test dataset. Of course, in practice, this is generally not the case. As the number of input variables and their ranges increase, the amount of data needed to adequately characterize the input space increases. Furthermore, these data need to contain enough variation to characterize important regions of the input space for future predictions.

A histogram comparing the significant wave height distributions of the AR Train, AR Test, and SR Test datasets is given in Figure 5.5. The normalized significant wave heights provided are those resulting from the data-driven model normalization approach, which was based on the maximum significant wave height seen in the AR Train dataset. Note that Figure 5.5 gives data densities rather than numbers of samples to allow more simple comparison between the AR Train dataset and the smaller test datasets. Also note, however, that the numbers above select bars correspond to numbers of samples, which are present wherever the number of binned samples fell below 100. For context, a lower number of binned samples in the AR Train dataset based on this single input variable increases the likelihood that test samples in this bin lack adequate representation for reliable predictions. As shown in Figure 5.5, the AR Train and Test datasets had higher densities at greater significant wave heights than the SR Test dataset, which was also implied by the wave statistics in Table 5.3. Although normalized significant wave height bins above 0.7 are relatively sparse in terms of AR Train samples, there are not any bins where test data are present without train data.

In contrast, Figure 5.6 shows a histogram of normalized mean wave period (T_m) that highlights bins below 0.3 and above 1.0 where test samples are present without

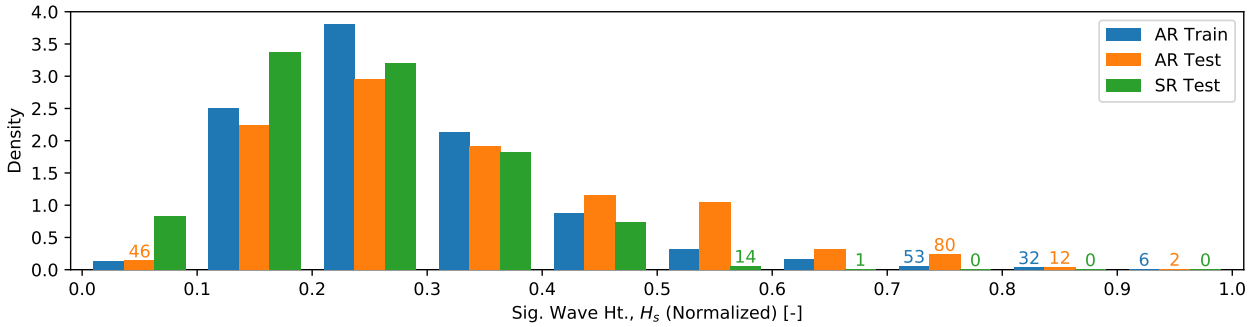

Figure 5.5: Histogram of normalized significant wave height comparing the AR Train, AR Test, and SR Test datasets. Although the bars reflect data density, the numbers listed above select bars are the total number of samples for a dataset in a given bin, and are included where these numbers falls below 100. The normalized significant wave heights are those resulting from the data-driven model normalization approach.

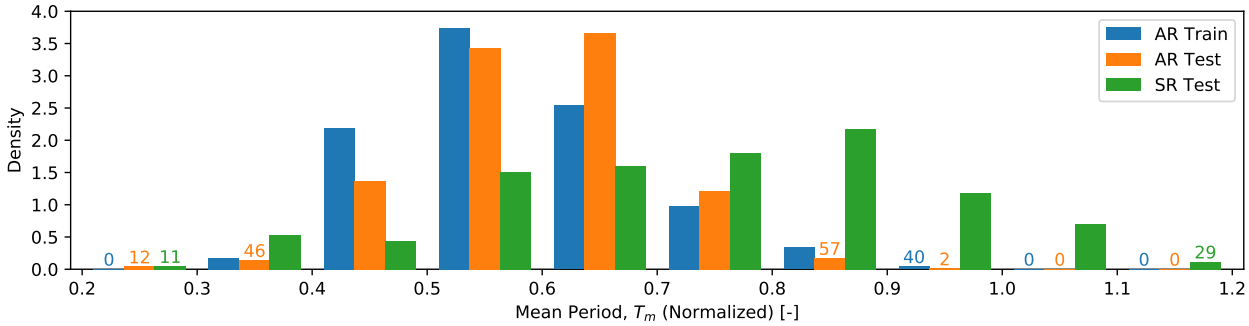

Figure 5.6: Histogram of normalized mean wave period (T_m) comparing the AR Train, AR Test, and SR Test datasets. Although the bars reflect data density, the numbers listed above select bars are the total number of samples for a dataset in a given bin, and are included where these numbers falls below 100. The normalized mean wave periods are those resulting from the data-driven model normalization approach.

corresponding train data. These normalized mean wave period values above 1.0 for SR Test were possible due to the data-driven model normalization approach, which was based on the maximum mean wave period seen in the AR Train dataset. The SR Test dataset from the *Sally Ride* had a higher density of samples with longer wave periods than the data from the *Neil Armstrong*, which was also implied in Table 5.3. This contrast is likely related to the expected wave profile differences between the Atlantic and Pacific Oceans. This shift between datasets adds another level of complexity and intrigue for application of these data-driven models, trained solely using data from the *Neil Armstrong*, to the *Sally Ride*. While the comparisons made between datasets

in this section were very limited relative to the number of input variables and possible conditions experienced, it is important to consider the role that dataset similarity (or dissimilarity) plays in the performance of these data-driven models and the potential benefits of including PBMPs to aid in train data-sparse regions of the input space.

5.2 Results and Discussion

The results discussed in this section were originally presented in *Schirmann et al.* (2020b) and *Schirmann et al.* (2021). The uncorrected PBMPs of heave, pitch, and roll standard deviations for the 3,384 AR Test windows, which were calculated using the initial twin framework, are plotted in Figure 5.7. The black lines have a slope of 1.0 and represent perfect agreement between measured and predicted values. Figure 5.7 shows reasonably strong agreement of uncorrected heave and pitch PBMPs with their respective measured values, especially in comparison to roll. As expected, these roll PBMPs were very poor, which allowed for the data-driven models to make significant improvements.

Figure 5.8 shows the 2,592 uncorrected PBMPs of the SR Test set. Comparing Figures 5.7 and 5.8, it is evident that the maximum SR Test dataset motions were less than those of AR Test. Furthermore, compared to AR Test, there are fewer roll

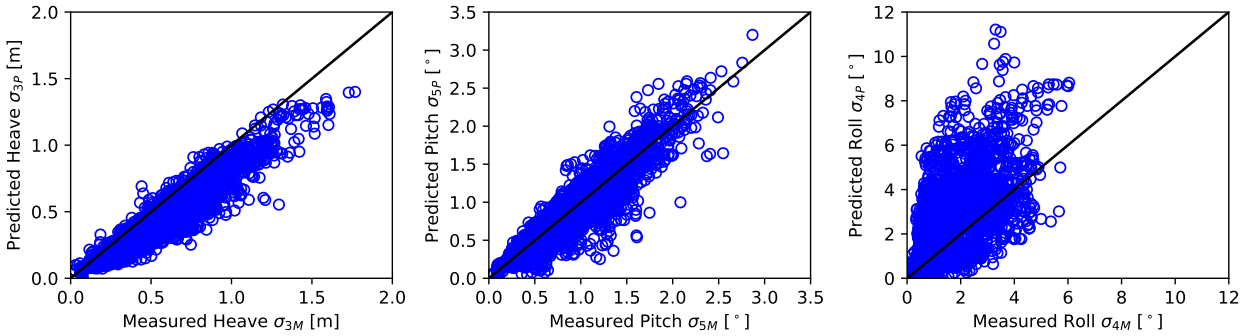

Figure 5.7: AR Test dataset uncorrected PBMPs of heave, pitch, and roll amplitudes calculated using the 2D strip-theory RAOs versus measured values. The black lines have slope 1.0 and represent perfect agreement between measured and predicted values for reference.

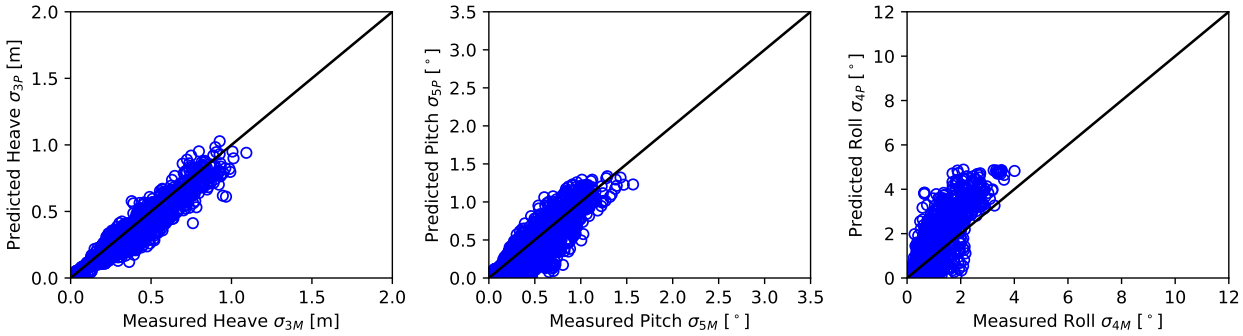

Figure 5.8: SR Test dataset uncorrected PBMPs of heave, pitch, and roll amplitudes calculated using the 2D strip-theory RAOs versus measured values. The black lines have slope 1.0 and represent perfect agreement between measured and predicted values for reference.

PBMPs that are drastically different than their respective measured values in the SR Test dataset. These lesser roll magnitudes and discrepancies may be linked to the lower significant wave heights or longer periods seen in the SR Test dataset, as discussed in Section 5.1.11. It is also possible that the actual and assumed mean loading conditions were more similar for the *Sally Ride* than they were for the *Neil Armstrong*, yielding more accurate PBMPs. Additionally, there could have been differences in the quality of measurements or subsequent filtering between datasets. For instance, it is possible that the measurement instruments and quality assessment for the *Sally Ride* removed more outlying (but not necessarily erroneous) measurements than the same combination for the *Neil Armstrong*. As such, the insights that can be drawn from direct comparison of AR Test and SR Test are limited. The PBMPs shown in Figures 5.7 and 5.8 serve as the baseline for comparisons in the following discussion.

5.2.1 R/V Neil Armstrong

This section details the results of data-driven model training and testing using the AR Train and AR Test datasets, respectively.

Table 5.9: Results of the 5-fold CV model selection processes for each of the four data-driven models. These selections were performed exclusively using the data in AR Train.

	RR		**NN**	
	Without PBMPs	With PBMPs	Without PBMPs	With PBMPs
Average CV Loss	0.004438	0.003544	0.003359	0.003273
Std. Dev. of CV Loss	0.001524	0.001431	0.001331	0.001463
Reg. Parameter, λ	0.155000	0.325000	0.000003	0.000030
HL Architecture	–	–	[40, 20, 10]	[40, 20, 10]

5.2.1.1 CV Selected Models and AR Train Performance

Table 5.9 summarizes the best models determined via 5-fold CV. Recall that the average CV loss, included in the first row of Table 5.9, was the performance metric used to determine the best parameters for each model from the user-specified candidates. The standard deviation of the CV loss from the five different groupings is also included for the best model, which shows that there were notable differences between model performances for each grouping. Intuitively, when random shuffling was applied in preliminary exploration of models, these CV standard deviations were significantly less due to the artificial similarity between subsets; nevertheless, as mentioned previously, there was no notable improvement in performance when applied to the test datasets. The average CV loss and standard deviations of CV loss are based on the normalized values used by the models.

The last two rows of Table 5.9 give the parameters that yielded the best models in terms of average CV loss. The HL architecture format [40, 20, 10] means that there were three HLs with 40, 20, and 10 neurons, respectively. While the best HL architectures with and without PBMPs were identical, note that the average CV losses resulting from significantly different HL structures were similar to these best model values; therefore, model performance on train data was not particularly sensitive to the HL architecture selected. For both data-driven model types, and especially for the NN where computational cost can be significant, it is important to recognize that

these best models were selected from a group of user-defined candidates rather than being the absolute best possible performance for a given dataset.

The heave, pitch, roll, and total MSE values of the PBMPs and data-driven approaches (after CV model selection) when applied to the entire AR Train dataset are included in Table 5.10. Recall that total MSE is the mean of the heave, pitch, and roll MSE values and was the metric used to train the models. While values in Table 5.10 demonstrate that the models learned from the train data, the insights that can truly be drawn from train data MSE are very limited. The AR Test performance evaluation discussed in Section 5.2.1.2 is far more significant.

Table 5.10: AR Train performance of the data-driven approaches with and without PBMPs, which are also included for reference. After the CV selection process outlined in Table 5.9 was completed for each model, the model was trained one final time on the entire AR Train dataset. The MSE values in this table are the result of those final training processes.

	Uncorrected	**RR MSE**		**NN MSE**	
	PBMPs' MSE	Without PBMPs	With PBMPs	Without PBMPs	With PBMPs
Heave	0.005955	0.002934	0.002319	0.001321	0.001460
Pitch	0.004441	0.004606	0.003255	0.001789	0.001954
Roll	0.075659	0.004373	0.003571	0.001548	0.001925
Total	0.028685	0.003971	0.003048	0.001553	0.001780

Table 5.11: AR Test performance of the data-driven approaches with and without PBMPs, which are also included for reference. The test data MSE is the best indicator of a model's real-world performance. For this reason, the superscripts are included to indicate the rank of each model in terms of minimizing MSE on the test set for a given DOF with (1) indicating the best model.

	Uncorrected	**RR MSE**		**NN MSE**	
	PBMPs' MSE	Without PBMPs	With PBMPs	Without PBMPs	With PBMPs
Heave	$0.007032^{(5)}$	$0.002696^{(4)}$	$0.002166^{(2)}$	$0.002292^{(3)}$	$0.002004^{(1)}$
Pitch	$0.004133^{(5)}$	$0.003603^{(4)}$	$0.002667^{(2)}$	$0.002760^{(3)}$	$0.002179^{(1)}$
Roll	$0.065751^{(5)}$	$0.008134^{(4)}$	$0.006372^{(3)}$	$0.003944^{(1)}$	$0.004108^{(2)}$
Total	$0.025639^{(5)}$	$0.004811^{(4)}$	$0.003735^{(3)}$	$0.002999^{(2)}$	$0.002764^{(1)}$

5.2.1.2 AR Test Performance

The heave, pitch, roll, and total MSE values of the PBMPs and data-driven approaches when applied to the AR Test dataset are included in Table 5.11. In contrast to train data values of Table 5.10, the test data MSE values in each DOF are the most indicative of a model's real-world performance when applied to an unseen dataset. Therefore, for each DOF in Table 5.11, the rank of each model in terms of test MSE minimization is included as a superscript, with (1) indicating the best model. The

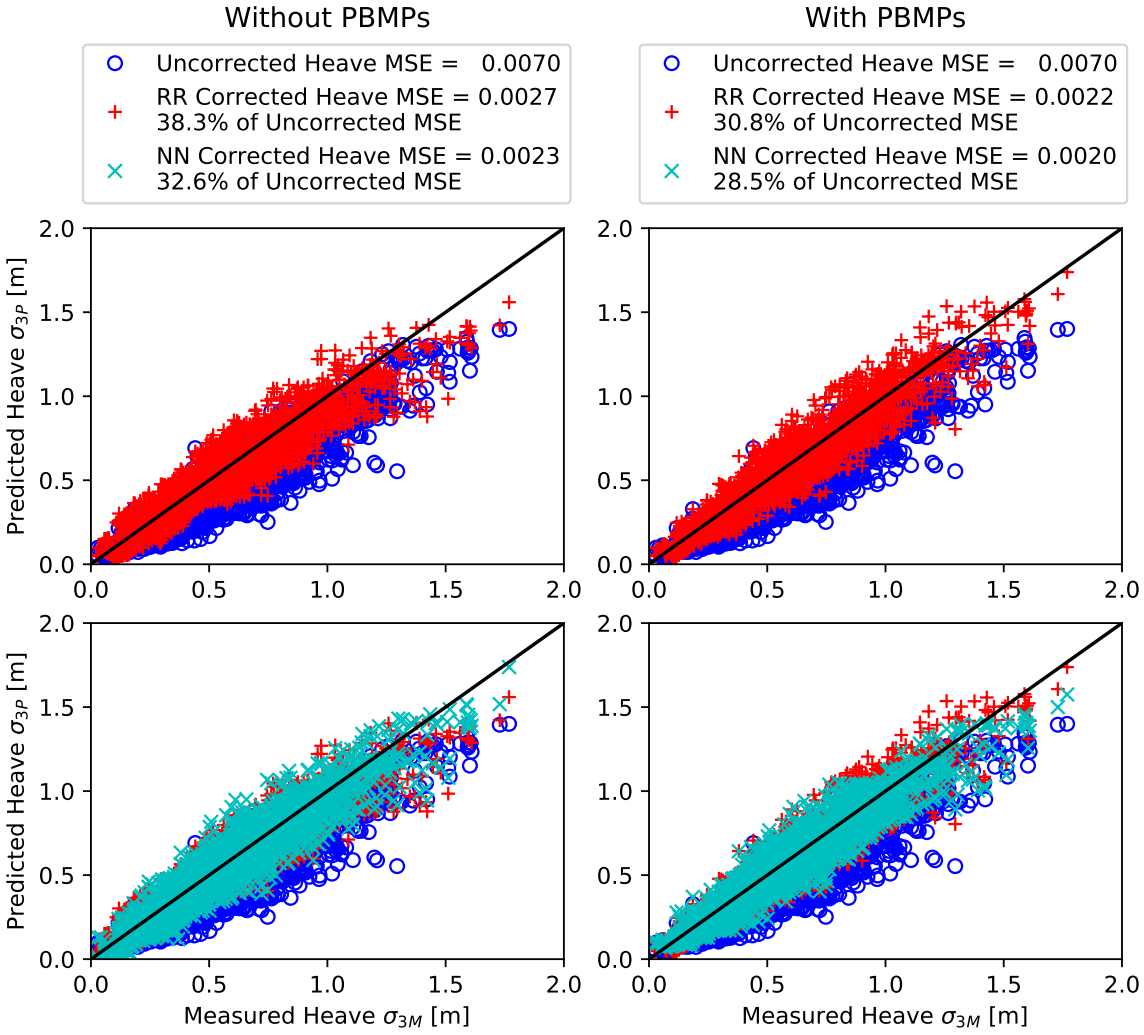

Figure 5.9: AR Test dataset uncorrected heave PBMPs and the RR and NN predicted heave amplitudes versus measured values. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference.

performances of these data-driven models on the test data are also demonstrated graphically for heave, pitch, and roll in Figures 5.9, 5.10, and 5.11, respectively. The top row of these figures shows the RR results overlaid on the uncorrected PBMPs (previously shown in Figure 5.7), and the bottom row shows the NN results overlaid on both the RR results and uncorrected PBMPs. As indicated, the left and right columns give the results without and with PBMPs included as input variables, respectively. The legends above each column give the approximate MSE of the normalized results for each approach, which match those presented in Table 5.11. The

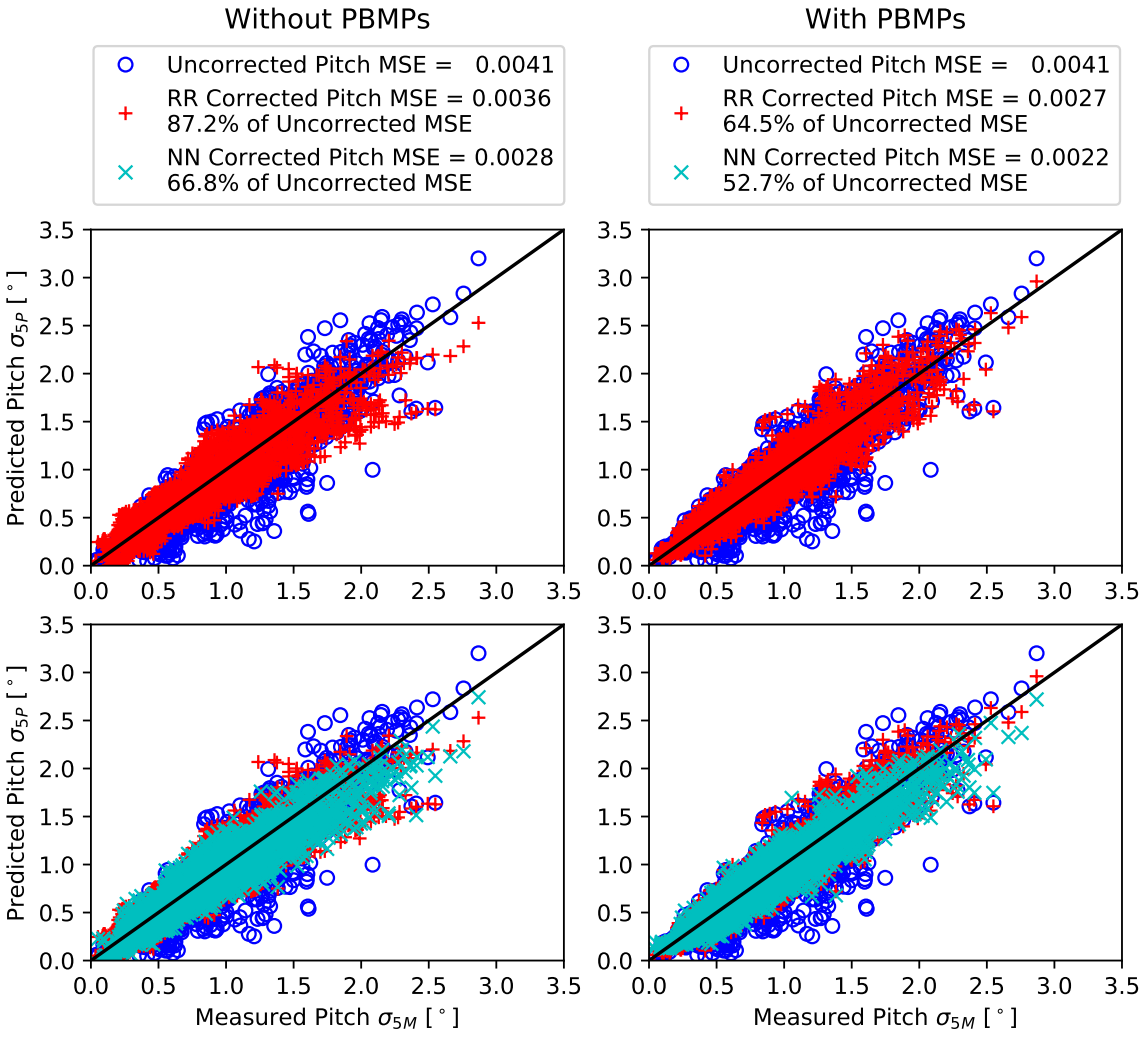

Figure 5.10: AR Test dataset uncorrected pitch PBMPs and the RR and NN predicted pitch amplitudes versus measured values. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference.

RR and NN legend entries also include their MSE value's relative percentage of the uncorrected PBMPs' MSE. These percentages highlight that for a given DOF and under the same PBMPs assumption (with or without PBMPs), the NN outperformed the RR approach in terms of MSE reduction for AR Test in all cases, which is consistent with the preliminary findings from Chapter IV. Once again, this consistently improved performance for the NN is largely attributed to its ability to model non-

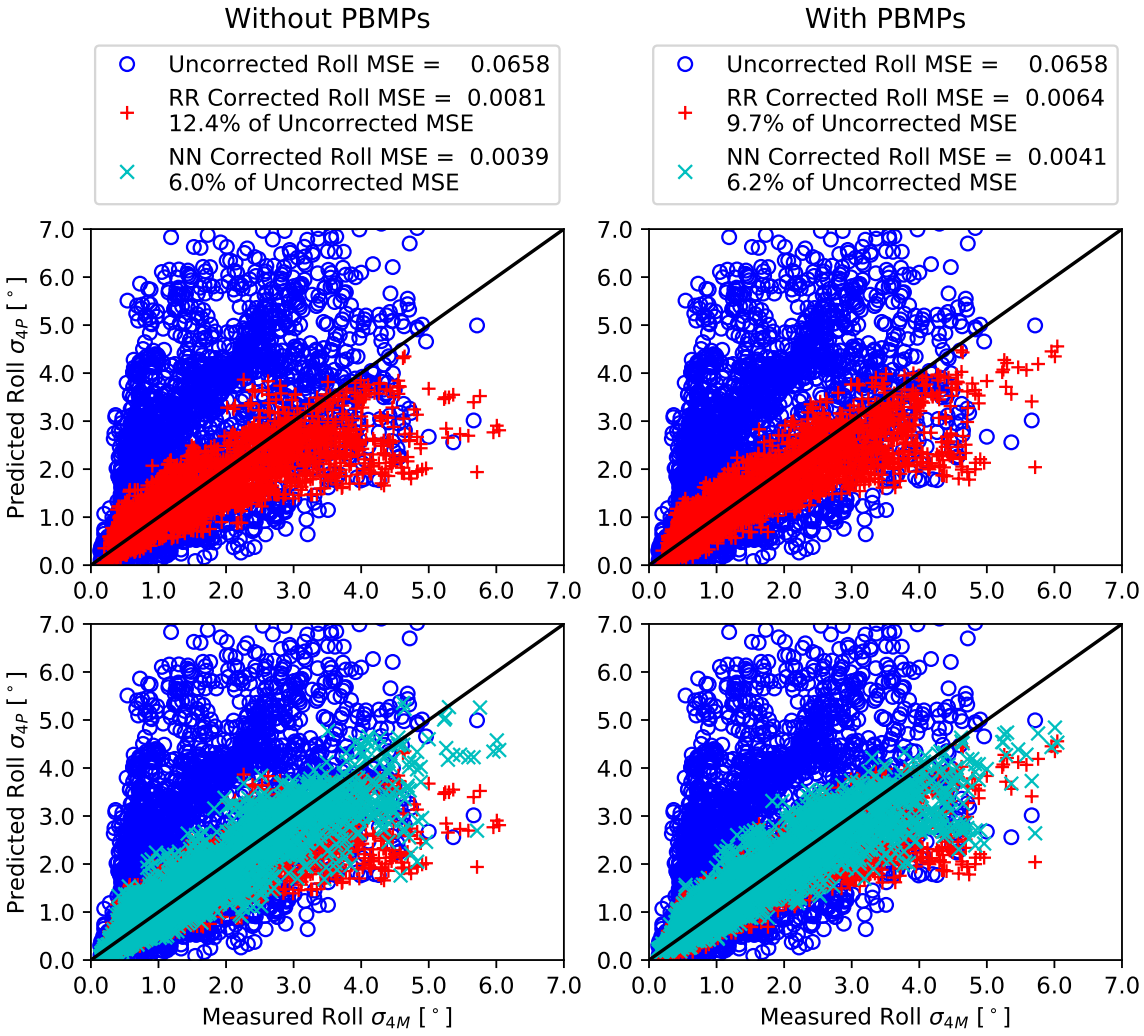

Figure 5.11: AR Test dataset uncorrected roll PBMPs and the RR and NN predicted roll amplitudes versus measured values. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference. Note that the roll PBMPs extend above the top of the plot as shown previously in Figure 5.7.

linear relationships; however, it is evident that the linear RR approach was still very effective for MSE reduction in all cases. Note that in Figure 5.11 for roll, the axes were limited to a value of 7.0° to highlight the RR and NN results, but the uncorrected PBMPs that were originally presented in Figure 5.7 extend to a value of nearly 12.0.

As indicated by the superscripts in Table 5.11, the NN model with PBMPs yielded the lowest total, heave, and pitch MSE values for the AR Test dataset. Therefore, the introduction of PBMPs as input variables added value and improved the heave and pitch predictions. Furthermore, for both heave and pitch, the second best approach was RR with PBMPs, which means that a linear approach (RR) outperformed a nonlinear approach (NN) when provided with PBMPs. This result was surprising and emphasizes the value that PBMPs can add to a model. This added value is largely attributed to the relatively strong accuracy of the uncorrected PBMPs, which was not the case for roll. As shown in Table 5.11, for roll, the NN without PBMPs was the best model and slightly outperformed the NN with PBMPs. This result means that no value was added by the PBMPs in the case of the nonlinear NN for roll. In the case of RR, the PBMPs did add value for all three DOFs, which is unsurprising given the limitations of a linear model. However, it should be noted that the improvement in this linear model may have resulted from the added DOFs provided by the PBMPs rather than the values themselves.

5.2.2 R/V Sally Ride Test Data

The efficacies of the *Neil Armstrong* trained NN and RR models with and without PBMPs when applied to the *Sally Ride* test dataset, SR Test, are detailed in Table 5.12. Once again, the SR Test MSE values are ranked, with (1) indicating the best model in each DOF. Additionally, the heave, pitch, and roll uncorrected PBMPs and RR and NN model predictions with PBMPs are plotted versus measured values in

Figure 5.12.

As shown in Table 5.12, for roll, the NN with PBMPs yielded the lowest MSE. This value added by roll PBMPs is contrary to the findings for AR Test, which may be due to the better roll PBMP quality for SR Test as discussed with regard to Figures 5.7 and 5.8. For heave and pitch, the top performer was the RR model with PBMPs, which contrasts the superior performance of the NN with PBMPs for the AR Test dataset as given in Table 5.11. The percentages given in Figure 5.12 clarify that for pitch, the performance of the NN with PBMPs was very similar to that of the RR with PBMPs, and both models significantly reduced MSE relative to the uncorrected PBMPs. On the contrary, for heave, Table 5.12 and Figure 5.12 show that the NN with PBMPs actually yielded a higher SR Test MSE than the uncorrected PBMPs. Note that the MSE of these uncorrected heave PBMPs was only slightly greater than MSE of the RR with PBMPs, meaning the best model made minimal improvements upon these PBMPs. While the differences in operating condition distributions of the AR Test and SR Test datasets, highlighted in Section 5.1.11, did not warrant a direct comparison of MSE values between datasets (Tables 5.11 and 5.12), note that the SR Test MSE values of the uncorrected heave, pitch, and roll PBMPs are all less than half of their respective AR Test values. This relationship is also evident visually

Table 5.12: SR Test results of the data-driven approaches with and without PBMPs, which are also included for reference. The test data MSE is the best indicator of a model's real-world performance. For this reason, the superscripts are included to indicate the rank of each model in terms of minimizing MSE on the test set for a given DOF with (1) indicating the best model.

	Uncorrected	**RR MSE**		**NN MSE**	
	PBMPs' MSE	Without PBMPs	With PBMPs	Without PBMPs	With PBMPs
Heave	$0.001383^{(2)}$	$0.001624^{(3)}$	$0.001332^{(1)}$	$0.002211^{(5)}$	$0.001650^{(4)}$
Pitch	$0.002148^{(4)}$	$0.002439^{(5)}$	$0.001432^{(1)}$	$0.002077^{(3)}$	$0.001473^{(2)}$
Roll	$0.012551^{(5)}$	$0.004031^{(4)}$	$0.003620^{(3)}$	$0.001920^{(2)}$	$0.001613^{(1)}$
Total	$0.005361^{(5)}$	$0.002698^{(4)}$	$0.002128^{(3)}$	$0.002069^{(2)}$	$0.001579^{(1)}$

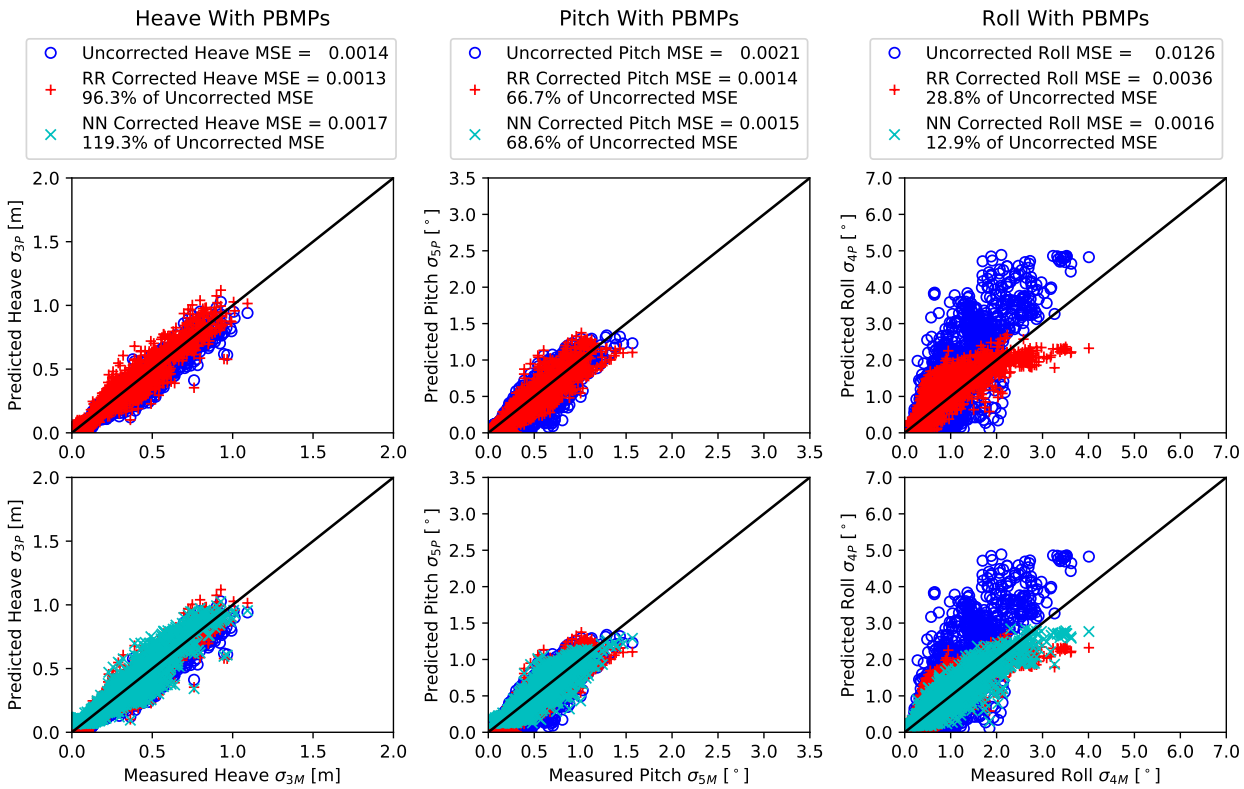

Figure 5.12: SR Test dataset uncorrected PBMPs and the amplitudes predicted by the RR and NN with PBMPs versus measured values for heave, pitch, and roll. The black lines have slope 1.0 and represent the line of perfect agreement between measured and predicted values for reference.

when comparing the uncorrected PBMPs given in Figures 5.7 and 5.8. For heave, specifically, the uncorrected PBMP's MSE of SR Test was four times smaller than that of AR Test. As such, the fact that the NN with PBMPs was slightly detrimental to vessel motion predictions may not be a major concern. Nevertheless, it is interesting that the RR with PBMPs outperformed the NN with PBMPs for the SR Test dataset. Additionally, Table 5.12 shows that for all three DOFs, the RR and NN models with PBMPs outperformed their counterparts without PBMPs, emphasizing the benefits of physics-based information.

5.2.3 Model Performance Comparisons

While Sections 5.2.1.2 and 5.2.2 demonstrated the benefits of including physics-based information for MSE reduction on each test dataset as a whole, there was not a discussion of the models' performances as a function of the measured motions or specific input variables. This section provides these more detailed model performance analyses.

5.2.3.1 Motion MSE Values versus Measured Motion Magnitudes

The true benefits of these data-driven models to a vessel owner or operator may be significantly greater when heave, pitch, or roll response magnitudes are large. Figure 5.13 shows heave, pitch, and roll MSE as a function of their respective normalized motion measurements. The three rows correspond to heave, pitch, and roll, and the three columns correspond to the AR Train, AR Test, and SR Test datasets. In each plot, a histogram of the number of samples in 10 different measured motion bins, identified by the right axis, is given. The data overlaid on these histograms are the MSE values calculated from the samples in each bin for a given model or the uncorrected PBMPs.

As discussed in Section 5.2.1.2, the test dataset MSE values are most indicative of a model's real-world performance. Furthermore, Sections 5.2.1.2 and 5.2.2 showed that in general, the RR and NN with PBMPs yielded similar or lower MSE values than their counterparts without PBMPs. As such, for each DOF (row), the y-axis bounds of Figure 5.13 were selected such that all test data MSE values (from AR Test and SR Test) were visible for the RR and NN with PBMPs. While it was deemed important to include the uncorrected PBMPs and the RR and NN without PBMPs to highlight bins where these models may have performed adequately or better than those with PBMPs, they are not the focal point of this discussion. Although train MSE is not a strong indicator of a model's real-world performance, the results for

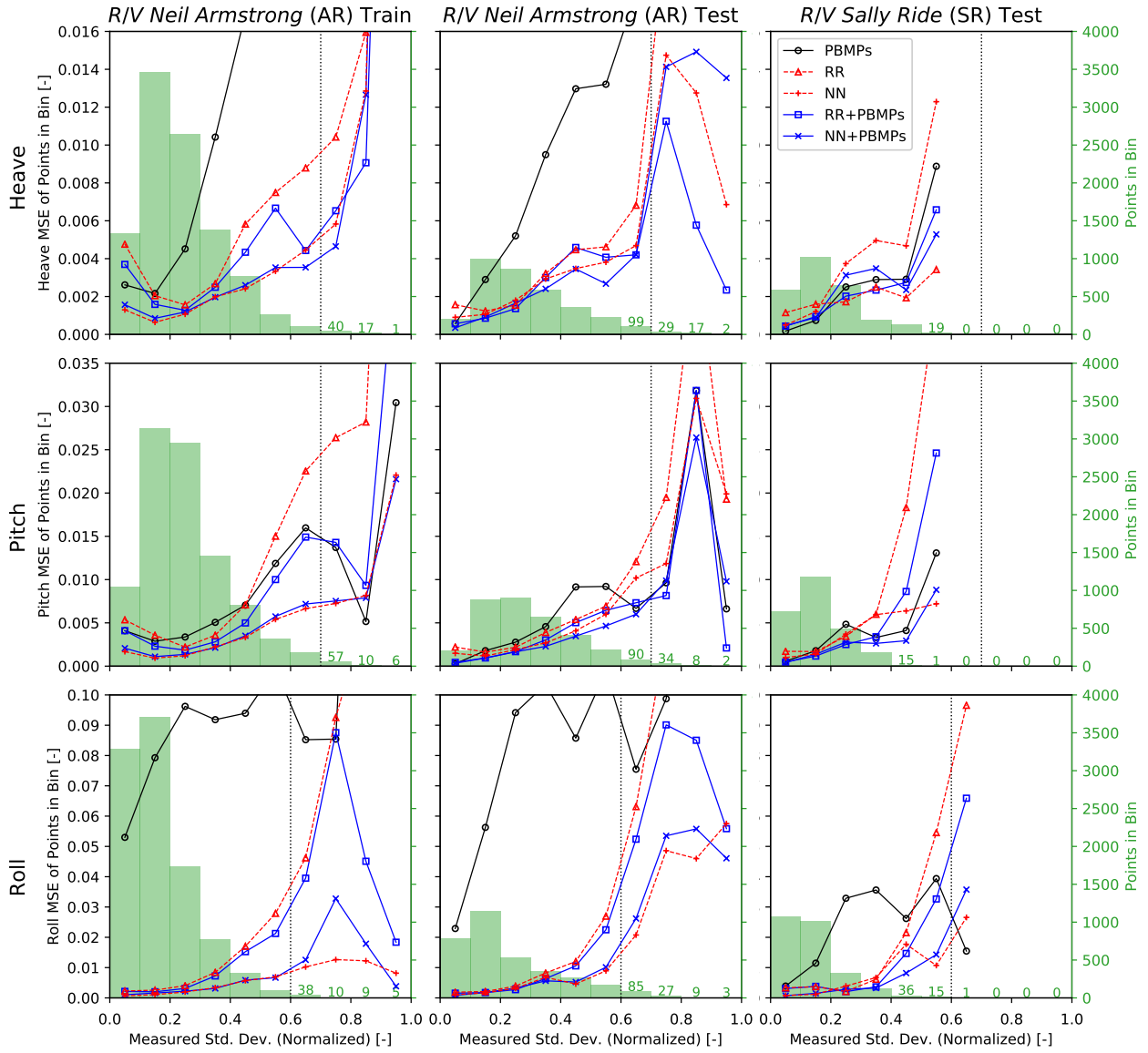

Figure 5.13: Plots of heave, pitch, and roll MSE as a function of their respective normalized motion measurements. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The MSE values, identified by the left axis, calculated from the samples in each bin for a given model or the uncorrected PBMPs, are overlaid on these histograms.

the AR Train dataset were included to indicate bins in which the models may have learned better than others. Similar to Figures 5.5 and 5.6, for magnitude bins with less than 100 samples, the number of samples is provided above the bar in Figure 5.13. Because train data sparsity can yield poor generalization, causing inadequate performance for a test dataset, the vertical dotted lines in each plot (for all three data sets) indicate the heave, pitch, or roll bin boundary where the number of AR Train samples dropped below 100. Note, however, that the selection of 100 was somewhat arbitrary and primarily motivated by plot legibility. Therefore, the insights provided by these dotted vertical lines are limited because model performance is tied to similarity between the train and test data for a given bin rather than the number of samples.

Figure 5.13 shows that typically, greater test MSE values occurred at greater motion magnitudes for all DOFs and predictive models. This increase in MSE was expected for several reasons. First, nonlinearities in vessel response typically grow as magnitudes increase, making accurate prediction more difficult. Second, proportional errors between predicted and measured values yield greater absolute errors as response magnitudes increase. Finally, the train data became more sparse as magnitudes increased, decreasing the likelihood of a test data sample having a similar sample in AR Train. Furthermore, Figure 5.13 shows that the differences in MSE values between models also typically grew as response magnitudes increased. This increase was also anticipated, especially when considering the expected advantages of models with PBMPs in sparse training data bins.

In Figure 5.13, for AR Test roll, the NNs with and without PBMPs yielded similar performance in most magnitude bins and outperformed the other models. As mentioned in the discussion of Table 5.11, this similarity may have resulted from the poor quality of the uncorrected roll PBMPs. Additionally, the advantages of the NNs over the RR models for roll prediction were likely due to the ability of NNs to model

nonlinearities. For SR Test, although the results of Table 5.12 implied that the NN with PBMPs notably outperformed the NN without PBMPs, Figure 5.13 suggests that the value added by PBMPs was only evident in a few bins. While the uncorrected roll PBMPs were seemingly better for SR Test than AR Test, the benefits of these PBMPs for roll were relatively insignificant. It is believed that improved roll PBMPs would improve the predictive capabilities of the NN with PBMPs, which is revisited in Chapter VIII.

For AR Test heave and pitch in Figure 5.13, the top performer for each magnitude bin was either the RR with PBMPs or NN with PBMPs. Furthermore, for some cases, such as AR Test pitch, the RR and NN without PBMPs performed significantly worse than the models with PBMPs and even the uncorrected PBMPs as magnitudes increased and, consequentially, train data were more sparse. This poor performance is an example cause of the mistrust that many engineers have in data-driven models that ignore physics. Note that there were a few select bins in which a model without PBMPs performed better than those with PBMPs (e.g., higher magnitudes for SR Test heave and pitch). Overall, for heave and pitch, the inclusion of PBMPs appears to have been beneficial. However, Figure 5.13 does suggest that any existent advantages of the NN with PBMPs over the RR with PBMPs were limited for heave and pitch.

5.2.3.2 Motion MSE Values versus Mean Wave Period

As shown in Figure 5.13, although some of the greater magnitude heave, pitch, and roll bins in AR Train had few data samples, none of the motion magnitudes in AR Test or SR Test fell outside of the AR Train bounds. As mentioned in Section 5.1.11, this was also the case for significant wave height, but not mean wave period (T_m). To examine the models' performances on test data outside of the AR Train bounds, Figure 5.14 follows the same general layout as Figure 5.13; however, all data in the plots of Figure 5.14 are binned by normalized mean wave period (T_m). Therefore,

for a given dataset (column) the histograms shown are identical for heave, pitch, and roll. The y-axis bounds were once again selected such that all AR Test and SR Test data samples for the NN and RR with PBMPs were included. Two dotted vertical lines indicating bins with less than 100 AR Train samples are included in the plots of Figure 5.14 because for mean wave period, test data extended to values above and below the AR Train range. Although AR Test is included in Figure 5.14 for reference, the samples lying outside of the mean wave period range of AR Train were limited. Therefore, SR Test is the focal point of the following discussion. Furthermore, it is difficult to draw conclusions from test data bins with very few samples (i.e., mean wave period bins 0.2-0.3 and 1.1-1.2 for SR Test) given the large number of other input variables. As such, the following discussion focuses on the normalized mean wave period bins 0.9-1.0 and 1.0-1.1. As shown by the histogram annotations, the 0.9-1.0 bin had just 40 AR Train samples, and the 1.0-1.1 bin had 0 AR Train samples.

For SR Test roll in the 0.9-1.0 bin of Figure 5.14, the NN with PBMPs performed notably better than any other model. For SR Test roll in the 1.0-1.1 bin, the NN with and without PBMPs were the top performers and yielded nearly identical MSE values. Therefore, for SR Test roll, the PBMPs may have been beneficial in combination with the 40 AR Train samples in the 0.9-1.0 bin, but did not have a notable effect in the 1.0-1.1 bin where 0 AR Train samples were available. For SR Test heave and pitch in the 0.9-1.0 bin, the RR with PBMPs yielded the lowest MSE. For this bin and heave, specifically, the RR with PBMPs was the only model with lower SR Test MSE than the uncorrected PBMPs, and the improvement was slight. For heave in the 1.0-1.1 bin, the uncorrected PBMPs produced notably lower MSE than any of the models. The uncorrected PBMPs were also the best performer for this bin in pitch, with the NN with PBMPs producing similar results. These results suggest that the inclusion of PBMPs was not definitively beneficial for extrapolated input variables when the *Neil Armstrong* trained models were applied to the *Sally Ride*.

This conclusion is not particularly surprising given the potential mean loading and operational environment differences between the two vessels. Perhaps a train dataset formed by data from both the *Neil Armstrong* and *Sally Ride* would be beneficial for future response predictions on both ships. If so, these data-driven models could be applied for fleet-level monitoring to support underway and deployment decisions.

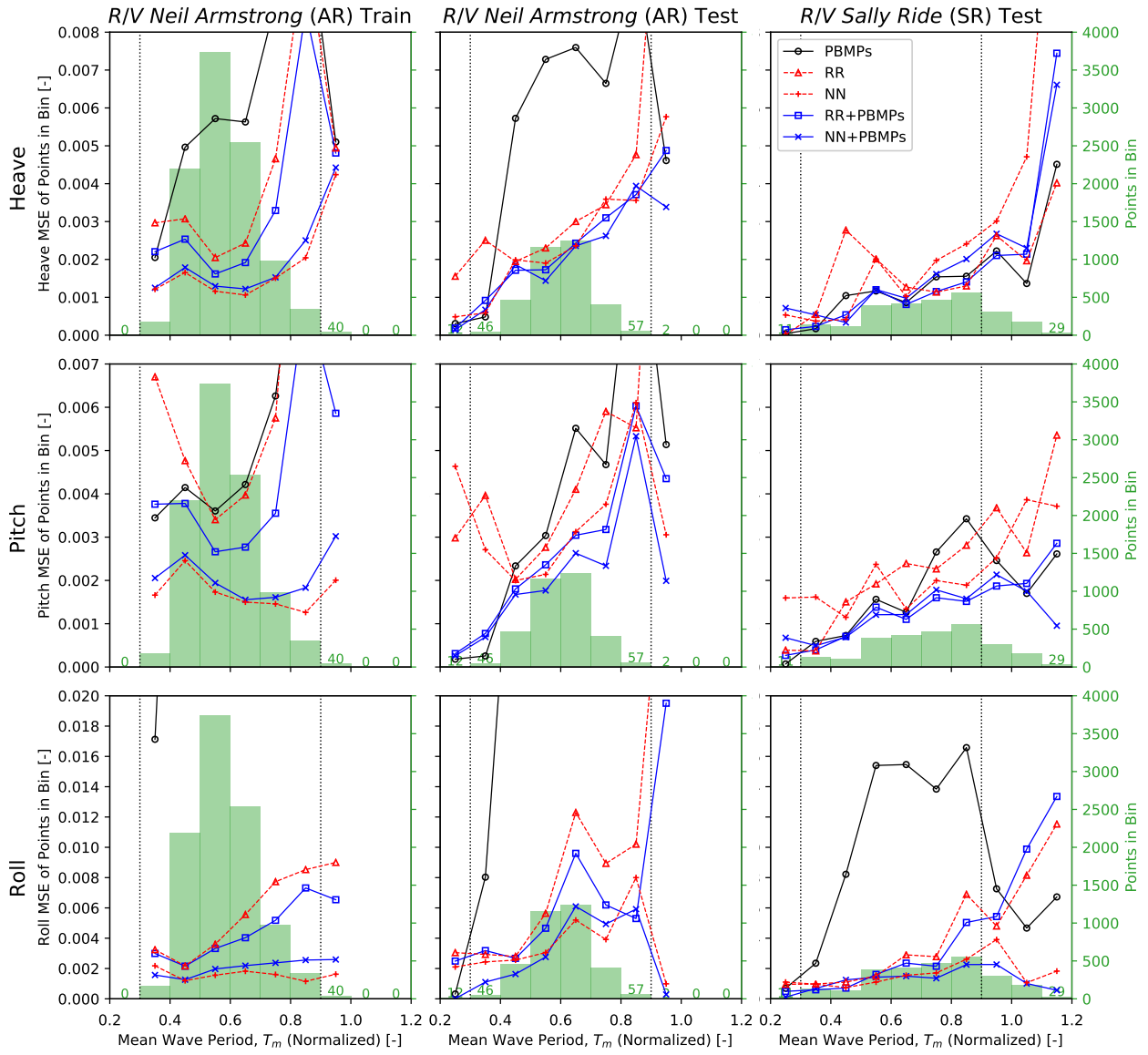

Figure 5.14: Plots of heave, pitch, and roll MSE as a function of normalized mean wave period (T_m). Each plot shows a histogram of the number of samples in 10 different mean wave period bins, identified by the right axis. The MSE values, identified by the left axis, calculated from the samples in each bin for a given model or the uncorrected PBMPs, are overlaid on these histograms.

5.2.4 AR40 Demonstration using Wave Forecast Data

As mentioned in Section 5.1.5, all of the CMEMS wave data used for AR Train, AR Test, and SR Test data were hindcast rather than forecast. In contrast, this section demonstrates the use of wave forecast data in the data-driven models, which emulates this technology's application in real-world operation of a ship for motion forecasting. As described in Section 5.1.5 and given in Table 5.3, AR40 yielded 132 30-minute windows for this analysis with five different wave forecast time horizons. Given the general benefits of including PBMPs discussed previously, the RR and NN models without PBMPs were omitted from the following.

Figure 5.15 shows the heave, pitch, and roll MSE values calculated from all 132 samples using wave data from the five different forecast time horizons defined in Section 5.1.5 and the hindcast wave data. Two plots with different zoom levels are included for roll due to the drastic improvement of both data-driven models upon the uncorrected PBMPs. For reference, the vertical dashed line in each plot identifies the transition from forecast to hindcast data, and the horizontal lines show the maximum MSE value over the six time horizons for each model.

In Figure 5.15, for heave and pitch, the NN with PBMPs yielded the lowest MSE for any given time horizon, followed closely by the RR with PBMPs in most instances. Furthermore, as expected, the use of hindcast data in the models yielded the lowest heave and pitch MSE for each model. With the exception of 72-48h for pitch, the heave and pitch MSE values for the two data-driven models decreased as the time horizons approached the time of interest (i.e., from left to right). For roll, these trends were not evident. In fact, the hindcast wave data yielded the greatest MSE for both the RR and NN with PBMPs. Note, however, that all the results shown in Figure 5.15 are extremely dependent upon the operating conditions and wave data available for this specific cruise; therefore, this discussion serves as a demonstration only, rather than an analysis from which broad conclusions regarding wave forecast

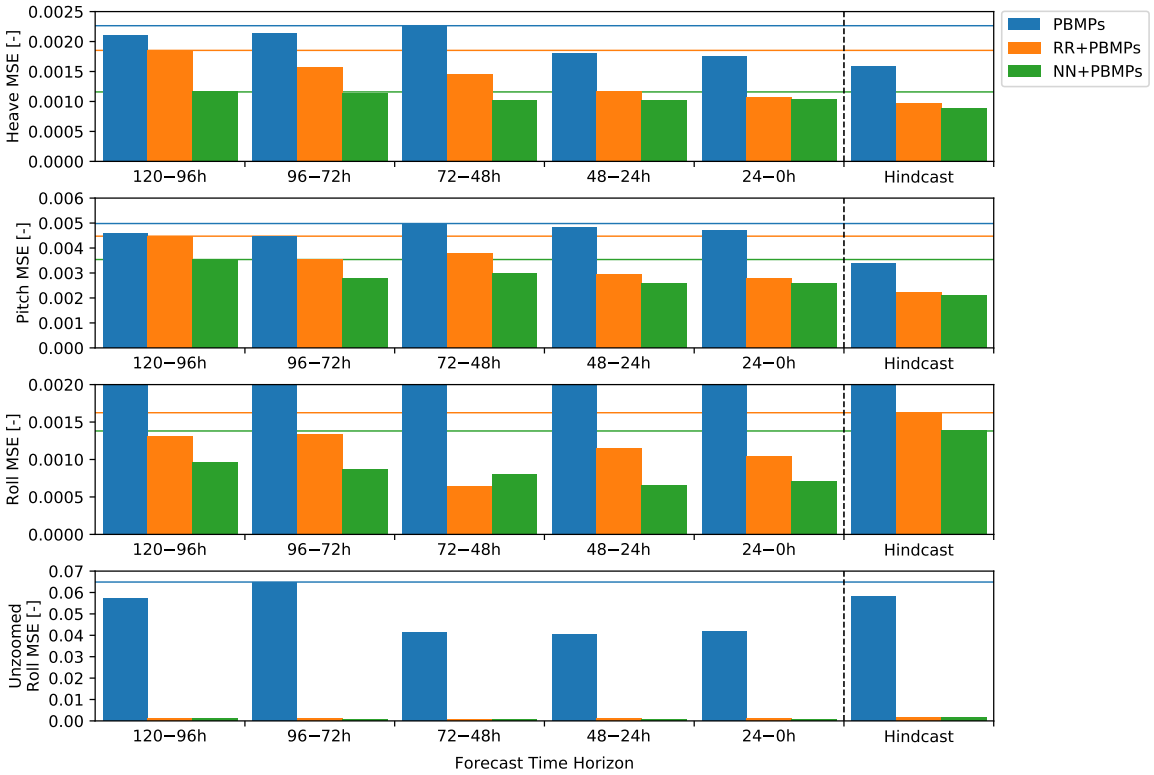

Figure 5.15: Heave, pitch, and roll MSE calculated from all 132 AR40 samples using the wave data from the five different forecast time horizons and the hindcast wave data.

reliability could be drawn.

This demonstration highlights the use cases outlined in Chapters I and II. In this scenario, given the vessel's expected location, speed, and heading, an operator could forecast how multidirectional wave conditions will affect the vessel up to five days in the future. These forecast motions would not only be informed by physics, but also by the vessel's past observations. As such, the operator could gain a better understanding of the conditions to come than that achieved using wave parameters alone. The operator could then continue to monitor these forecasts as the time of interest approached, as demonstrated in Figure 5.15, and adjust course and/or the vessel's estimated time of arrival, if necessary.

5.3 Research Vessel Motion Prediction Conclusions

The results described in this chapter, originally presented in *Schirmann et al.* (2020b) and *Schirmann et al.* (2021), demonstrated the application of linear RR and nonlinear NN models for vessel heave, pitch, and roll prediction using multidirectional wave data and over 16,000 30-minute measurement windows from two operational sister ships. The benefits of retaining physics-based information in data-driven models were explicitly demonstrated through comparison of model performance with and without inclusion of PBMPs as input variables. The models were trained and tested using data from the *R/V Neil Armstrong*. While both the RR and NN approaches effectively reduced the AR Test MSE relative to the uncorrected PBMPs, under the same PBMPs assumption (with or without PBMPs), the NNs outperformed the RRs. As in Chapter IV, this improved performance was attributed to the ability of the NNs to model nonlinear relationships between input variables. For AR Test heave and pitch, which had PBMPs of much higher quality than roll, the inclusion of PBMPs benefited model performance. In fact, the second most effective model behind the NN with PBMPs was the RR with PBMPs, showing that a linear model (RR) with PBMPs could outperform a nonlinear model (NN) without PBMPs. For roll, the low-quality PBMPs did not show any benefit for the NN and limited benefit for the RR approach.

These *Neil Armstrong* trained models were then applied to a test dataset from the *R/V Sally Ride* to test their versatility. In this case, the best approaches for heave and pitch were the RR models with PBMPs, and the best model for roll was the NN with PBMPs. Excluding the NN with PBMPs for heave, all models that included PBMPs reduced the heave, pitch, and roll MSE values of the test dataset relative to the uncorrected PBMPs. The exception for heave may have been due to the relatively high quality of the uncorrected heave PBMPs for the SR Test dataset. Although the relative performance comparisons between models for SR Test did not

perfectly match the findings from AR Test, the results did show that in each DOF, the RR or NN with PBMPs outperformed their respective counterparts without PBMPs. This result highlights the potential benefits of including physics-based information in data-driven models.

A more in-depth analysis of the motions MSE values as a function of measured motion magnitudes further supported that quality PBMPs, as in the case of heave and pitch, typically improved model performance. Additionally, these comparison demonstrated that in some cases, a lack of PBMPs can cause data-driven models to perform more poorly than the uncorrected PBMPs. This finding reinforces the concerns of some regarding the reliability of data-driven models that abandon physics. An additional analysis of motions MSE values as a function of mean period in train data sparse regions of the input space did not show definitive benefits from including PBMPs when the *Neil Armstrong* trained model was applied to *Sally Ride* data. Nevertheless, training RR and NN models with PBMPs using data from both ships may prove more effective for future response predictions on both ships and benefit fleet-level monitoring efforts.

A demonstration of the RR and NN models with PBMPs using forecast, rather than hindcast, wave data was performed for a single cruise of the *Neil Armstrong*. The potential use of these models for real-world operational guidance was highlighted through comparison of model performance ranging from a five-day forecast to a hindcast analysis. As discussed in Chapters I and II, providing vessel operators with these motion predictions may furnish a better understanding of a multidirectional wave forecast's meaning, leading to safer and more efficient vessel operation.

Using data from over 16,000 30-minute windows measured aboard two operational sister ships, this study demonstrated the efficacy of data-driven models for vessel motion prediction using wave model data. Furthermore, the results in this chapter highlighted the importance of including physics-based information, even given an

abundance of train data.

CHAPTER VI

Shared-Layer NN for Research Vessel Motion Predictions

This chapter builds upon the data-driven, physics-informed models presented in Chapter V for vessel heave, pitch, and roll prediction using multidirectional wave model data. A shared-layer neural network (SLNN) with a physics-informed structure was constructed, trained, and tested using data from the two sister vessels. This structure was designed to enforce physics-based behavior and improve the versatility of the predictive framework such that it could be used with a varying number of directional wave partitions and perform reliable wave data interpolation in space and time. More specifically, the developed SLNN model structure addressed the following shortcomings of the traditional NNs constructed, trained and tested in Chapter V:

1. Although different wave partition types (e.g., wind seas, primary swell, secondary swell) should generally be governed by the same physical principles, the traditional NN structure does not enforce this similarity between input variables for different wave partitions.

2. The traditional NN structure grows with the addition of wave partitions, making it less feasible for reliable response prediction with wave data sources that have a greater number of directional partitions than CMEMS.

3. The traditional NN structure is not readily adaptable for wave data sources with a varying number of wave partitions, such as the NWW3 Production Hindcast used in Chapter III (*NOAA National Centers for Environmental Prediction.*, 2019).

4. The wave data interpolation approach used to prepare NN input variables was flawed because CMEMS' identification of the three wave partitions was not necessarily consistent between wave data output times and locations. The use of wave data sources with more wave partitions makes the need to address this shortcoming even more significant.

These shortcomings and the ways in which the developed SLNN structure addresses each are described in the following paragraphs.

The best-performing RR and NN models from Chapter V incorporated physics-based information in the form of computationally efficient PBMPs of the vessel's motion amplitudes. These PBMPs were generated using each of the directional CMEMS wave partitions (WW, $SW1$, and $SW2$) and the initial twin framework. Recall that along with the vessel's speed and CMEMS wave parameters, these calculated heave, pitch, and roll amplitudes due to each wave partition were injected into the models as nine input variables. The total heave, pitch, and roll amplitudes resulting from all wave partitions were also provided as three additional input variables. Although the RR and NN models incorporated physics-based information via these 12 PBMP variables, the models did not necessarily handle the heave, pitch, and roll responses from the three different wave partitions equivalently. For example, in the CMEMS data structure, the primary swell was defined as the swell partition with the greatest amount of energy (i.e, significant wave height). From a physics standpoint, if the primary and secondary swells were nearly identical, the data-driven model output should be influenced nearly identically by both; however, the model structure did not guarantee this behavior. The network's handling of wave parameters and associated

PBMPs could vary between wave partitions. Furthermore, while the influence of swell and wind wave partitions on vessel motions are likely not identical, most of the underlying physical principles that govern these interactions are the same, but the models' structures did not enforce this similarity. As a researcher or engineer constructing these types of networks, one can only hope that similar physics are learned from the train dataset and, given the black box nature of NNs, it is difficult or impossible to definitively verify this. The approach developed here enforces this similarity using the network's structure.

Despite the concerning lack of enforced physics-sharing between wave partitions for the CMEMS-based NNs of Chapter V, the relatively small number of wave partitions with defined types (wind waves or swell) made the traditional, feed-forward NN structure feasible. In contrast, given a wave data source with a much larger number of wave partitions, possibly even one with discretely defined 2D wave spectra, the need to enforce physics-sharing and maintain a tractable input variable space becomes significantly more critical. For example, consider the six variables (comprised of wave parameters and PBMPs) specific to each CMEMS wave partition in Chapter V; addition of just one wave partition would increase the total number of input variables to the network by 20%. Also consider a wave data source with a varying number of wave partitions, such as the NWW3 Production Hindcast, which often includes significantly more wave partitions than CMEMS' three. If a maximum of 10 wave partitions were defined for samples in corresponding train and test datasets, but only one sample in the train dataset actually had this many partitions, the network's response to the tenth wave partition for test samples would be based solely on this single train sample. Worse, if a tenth wave partition was present in one or more test samples and zero samples of the train dataset, the model could predict completely erroneous motions for these test samples. In contrast, a model structure that shares physics between wave partitions could apply learned behavior from any

wave partition of any train data sample to this tenth wave partition. To demonstrate this scenario, in addition to CMEMS, two unique data sources with varying numbers of wave partitions were employed in this chapter.

As mentioned in Chapter V, the wave data interpolation approach was imperfect because CMEMS' identification of the three wave partitions was not necessarily consistent between wave data output times and locations. For instance, primary and secondary swell could swap between two interpolation points, which could yield unreliable interpolated values, especially for the case of wave direction parameters. While these instances were likely infrequent in the study of Chapter V, this study seeks to address wave data interpolation in a more robust manner through the developed model structure.

As mentioned above, the goal of this work was to construct a data-driven model structure that shares physics between directional wave partitions through shared network layers. The developed SLNN structure also leverages this shared-layer approach for more robust wave data interpolation in space and time than the study of Chapter V. An added benefit of the SLNN described in the following discussion is that it allows the use of any number of wave partitions, rather than just the three available through CMEMS. Recall that the NWW3 multidirectional wave data employed in the wave data source selection study of Chapter III were defined using a flexible number of wave partitions. To demonstrate this approach, SLNN construction, training, and evaluation was performed using three different wave data sources, each with a unique wave data format and number of wave partitions. This study was conducted using the same onboard measurement dataset from Chapter V of approximately 16,000 30-minute windows from the sister vessels *R/V Sally Ride* and *R/V Neil Armstrong*.

One of the wave data sources employed in this study was the same as used in Chapter V (CMEMS). In this study, CMEMS wave data were used in two different model constructions. The first method followed the full SLNN construction detailed

in this chapter. To draw similarities between the present study and the study of Chapter V, the second method was very similar to the previous NN, with no explicit physics-sharing between the three wave partitions; however, the NN employed in this study did incorporate the more robust wave data interpolation approach. As in the wave data source selection study of Chapter III, NWW3 multidirectional wave data, which had a flexible number of wave partitions, were used as the second wave data source for SLNN construction. Finally, ERA5 2D wave spectra were acquired from the Copernicus Climate Change Service Climate Data Store (CDS) as another unique data source for demonstration of the SLNN. Note, as in the wave data source selection study of Chapter III, the goal of this study was not to establish a hierarchy of wave data sources or comment on their validity. The purpose of considering multiple wave data sources with unique variables and data structures was to demonstrate the versatility of the developed SLNN approach.

6.1 Methodology

This section details SLNN model construction, training, and testing performed using data from three different wave data sources and the datasets discussed in Chapter V from the sister vessels *R/V Neil Armstrong* and *R/V Sally Ride*.

6.1.1 Wave Data Sources

As mentioned previously, the three hindcast wave data sources employed in this chapter were CMEMS, NWW3, and CDS ERA5 2D wave spectra, which are described in *Fernandez and Aouf* (2018), *NOAA National Centers for Environmental Prediction.* (2019), and *Copernicus Climate Change Service [C3S]* (2021); *Hersbach et al.* (2020), respectively. The NWW3 data source corresponded to the NWW3 Glo 30 min source of Chapter III. As seen in previous chapters, CMEMS and NWW3 were provided as characteristic wave parameters for a given time and location. In

contrast, the CDS wave spectra were provided as 2D spectral densities for 30 wave frequencies and 24 evenly spaced directions on [0°, 360°). In this study, each of the 24 CDS directions were treated as individual wave partitions, with the resultant 1D spectral density for each directional wave partition used in the initial twin framework for response prediction, as discussed in Section 6.1.2. However, in this study, these defined wave spectra were not used directly as input to the SLNN model structure, but rather the spectral data were used to calculate characteristic parameters matching those of CMEMS. Note that in future studies, the 30 spectral densities for each directional wave partition could be used as input variables in place of characteristic parameters.

Table 6.1: Summary of the three wave data sources and the number of 30-minute onboard measurement windows available for each dataset.

Wave Data Source	Lat/Lon Resolution	Time Step	Max WPs Observed	Provided Wave Data Format	AR Train Windows	AR Test Windows	SR Test Windows
CMEMS	1/12°	3 h	3	Characteristic Parameters	10,034	3,387	2,622
CDS	1/2°	1 h	24	Discrete Spectral Densities	10,034	3,387	2,622
NWW3	1°	1 h	13	Characteristic Parameters	10,034	2,256	2,431

Table 6.1 summarizes the three wave data sources, including their latitude/longitude and time resolutions, which varied notably. For instance, CMEMS had the highest spatial resolution, but CDS and NWW3 had higher temporal resolution. Furthermore, the maximum number of wave partitions observed across all considered training windows were significantly different between wave data sources. Therefore, as mentioned previously, the purpose of this study was not to compare the wave data sources, but to demonstrate the versatility of the SLNN's structure. Note that the NWW3 latitude/longitude resolution of Table 6.1 is 1°, which contrasts the 0.5° specification from Chapter III. As mentioned in Chapter III, there were different masking patterns in the NWW3 datasets to reduce file sizes. As such, even though the NWW3 data were generated using a 0.5° grid, the grid was masked to the 1° resolution applicable

for interpolation in this study.

Table 6.1 includes the number of 30-minute windows in the AR Train, AR Test, and SR Test datasets available for each wave data source. Note that the number of windows in each dataset for CMEMS increased slightly from Chapter V (e.g., 10,034 windows for AR Train rather than the 10,008 given in Table 5.4). This change was due to a shift in the handling of wave data interpolation where and when one or more data points provided by a wave data source and considered for interpolation were masked, which is described in Section 6.1.3. Also note that for the NWW3 dataset, data were unavailable beyond May 2019; therefore, the AR Test and SR Test datasets had fewer points than their counterparts for CMEMS and CDS. This discrepancy is another reason that direct comparison between results for the different wave data sources would be inappropriate in this study.

The characteristic wave parameters for each wave data source could be divided into two categories: global variables, which describe the wave conditions due to all wave partitions (e.g., total significant wave height), and local variables, which describe a single directional wave partition (e.g., primary swell significant wave height). For a given time and location, these two categories can be viewed as two vectors of wave data variables: global, $\mathbf{g}_i$, and local, $\boldsymbol{\ell}_{ij}$. Note that the subscript i refers to each of the eight time and location "corners" used in wave data interpolation, similar to Chapter V and detailed in Section 6.1.3. For local variable vector $\boldsymbol{\ell}_{ij}$, the subscript j indicates a singular wave partition. The wave data variables stored in $\mathbf{g}_i$ and $\boldsymbol{\ell}_{ij}$ for each wave data source are provided in Tables 6.2 and 6.3, respectively. Note that the wave direction parameters stored in both $\mathbf{g_i}$ and $\boldsymbol{\ell}_{ij}$ are adjusted to be relative to the vessel's heading.

In Tables 6.2 and 6.3, note the use of wave elevation variance rather than significant wave height. This contrast with previous chapters was employed for consistency with the SLNN structure's output variables, motion variances, which are detailed in

Table 6.2: The global wave data variables stored in $\mathbf{g}_i$ for each wave data source. CDS is marked with * to highlight that its global variables were derived from the provided 2D wave spectra to match those of CMEMS.

Variable	CMEMS	CDS*	NWW3	Units	Description
σ_η^2	✓	✓	✓	m^2	Wave Elevation Variance (a.k.a. zeroth spectral moment, $m_{\eta 0}$)
T_p	✓	✓	✓	s	Spectral Peak Wave Period
T_{01}	✓	✓		s	Mean Wave Period (m_0 and m_1)
T_{02}	✓	✓		s	Mean Wave Period (m_0 and m_2)
$\overline{\theta_{Rel}}$	✓	✓		deg	Mean Wave Direction (Relative to Ship Heading)
$\theta_{p,Rel}$	✓	✓		deg	Wave Direction at Spectral Peak (Relative to Ship Heading)
WSF			✓	–	Wind Sea Fraction on [0,1] (0 for pure swell, 1 for pure wind seas)
$\lambda(T_p)$			✓	m	Wave Length at Peak Period
$\overline{\theta_{Rel}(T_p)}$			✓	deg	Mean Wave Direction at Peak Period (Relative to Ship Heading)
$\Delta\theta$			✓	deg	Directional Spectral Width

Table 6.3: The local wave data variables for wave partition j stored in $\boldsymbol{\ell}_{ij}$ for each wave data source. CDS is marked with * to highlight that its local variables were derived from the provided 2D wave spectra to match those of CMEMS.

Variable	CMEMS	CDS*	NWW3	Units	Description
σ_η^2	✓	✓	✓	m^2	Wave Elevation Variance (a.k.a. zeroth spectral moment, $m_{\eta 0}$)
T_p			✓	s	Spectral Peak Wave Period
T_{01}	✓	✓		s	Mean Wave Period (m_0 and m_1)
$\overline{\theta_{Rel}}$	✓	✓		deg	Mean Wave Direction (Relative to Ship Heading)
Q	✓			–	CMEMS Wave Partition Type (1 for WW, 0 for $SW1$ or $SW2$)
WSF			✓	–	Wind Sea Fraction on [0,1] (0 for pure swell, 1 for pure wind seas)
$\lambda(T_p)$			✓	m	Wave Length at Peak Period
$\overline{\theta_{Rel}(T_p)}$			✓	deg	Mean Wave Direction at Peak Period (Relative to Ship Heading)
$\Delta\theta$			✓	deg	Directional Spectral Width

Section 6.1.4. In Table 6.3, in addition to the CMEMS variables from Chapter V, note the addition of indicator variable Q to distinguish between the wind sea and swell partitions, but not between primary or secondary swell. While wind sea and swell partitions should be governed by similar physics, this variable within the SLNN structure enforced that different swell partitions were governed by identical physics. The CDS local variable vector did not contain this indicator variable Q because wind sea or swell identification was not provided or derived for the 24 directional wave partitions. The use of $\mathbf{g}_i$ and $\boldsymbol{\ell}_{ij}$ in the SLNN and the addition of corresponding global and local PBMPs as model input are described in Section 6.1.4.

6.1.2 Physics-Based Model Predictions (PBMPs)

As in previous chapters, the initial twin framework was used to generate response predictions using each wave data source's directional wave partitions (characterized by the variables in $\boldsymbol{\ell}_{ij}$). For CMEMS and NWW3, Bretschneider spectra were constructed from each wave partition's characteristic wave parameters because defined spectral shapes were unavailable. In contrast, as mentioned in Section 6.1.1, the defined 1D wave spectra from the 24 directional wave partitions were used for the CDS wave data source. As in Chapter V, no directional spreading was applied for any wave data source or directional partition. Note that unlike the study of Chapter V, heave, pitch, and roll PBMPs were generated for each of the eight wave data corners in latitude, longitude and time space rather than using a single interpolated wave data point. For presentation of these PBMPs in comparison to the data-driven models, the following interpolation approach was applied to combine the PBMPs into a single heave, pitch, and roll response. The SLNN's interpolation scheme is described in the following section.

6.1.3 Interpolation Approach

In Chapter V and this study, each selected 30-minute window was a single sample with an associated waypoint, defined by a mean time ($\overline{t}$), latitude ($\overline{\text{lat}}$), and longitude ($\overline{\text{lon}}$). Each waypoint, ($\overline{t}, \overline{\text{lat}}, \overline{\text{lon}}$), fell between points (or on a single point) contained in the three-dimensional (time, latitude, and longitude) wave data output grid. The time, latitude, and longitude increments that defined this wave data output grid were denoted Δt, Δlat, and Δlon, respectively, and correspond to those presented in Table 6.1 for each wave data source. The specific time, latitude, and longitude grid intervals (e.g., $[t_0, (t_0 + \Delta t)]$) containing each sample's waypoint were identified as shown in

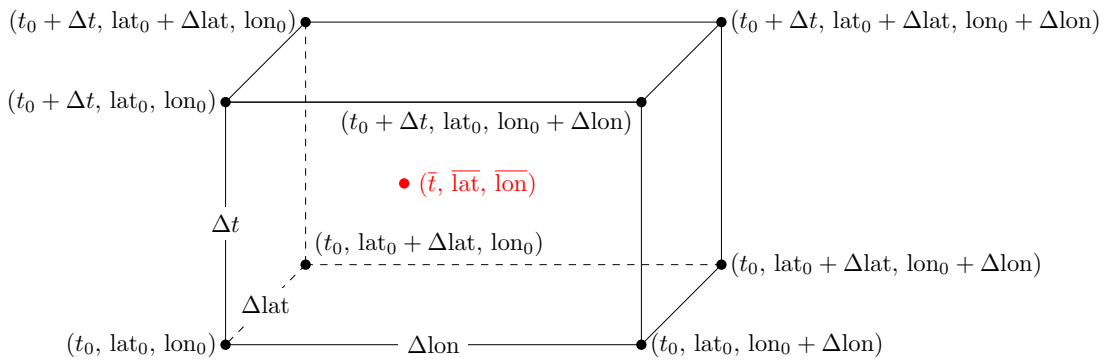

Figure 6.1: Interpolation prism.

Equation 6.1.

$$\begin{aligned} t_0 &\leq \bar{t} \leq (t_0 + \Delta t) \\ \mathrm{lat}_0 &\leq \overline{\mathrm{lat}} \leq (\mathrm{lat}_0 + \Delta\mathrm{lat}) \\ \mathrm{lon}_0 &\leq \overline{\mathrm{lon}} \leq (\mathrm{lon}_0 + \Delta\mathrm{lon}) \end{aligned} \tag{6.1}$$

Combination of the bounds identified in Equation 6.1 yields eight (2^3) total points from the wave data output grid that are relevant for linear interpolation at $(\bar{t}, \overline{\mathrm{lat}}, \overline{\mathrm{lon}})$. Geometrically, these eight points can be viewed as "corners" of a rectangular prism containing $(\bar{t}, \overline{\mathrm{lat}}, \overline{\mathrm{lon}})$, as depicted in Figure 6.1. As alluded to in discussion of $\mathbf{g}_i$ and $\boldsymbol{\ell}_{ij}$, each of these corners was identified by an index, i, as detailed in Table 6.4.

Table 6.4: Coordinates of the rectangular prism's eight corners for a given waypoint in time, latitude, and longitude space. The relevant prism from the wave data output grid contains the waypoint, $(\bar{t}, \overline{\mathrm{lat}}, \overline{\mathrm{lon}})$, as demonstrated in Equation 6.1. These eight (2^3) values were collectively necessary for linear interpolation in three dimensions.

i	t_i	lat_i	lon_i
1	t_0	lat_0	lon_0
2	t_0	lat_0	$\mathrm{lon}_0 + \Delta\mathrm{lon}$
3	t_0	$\mathrm{lat}_0 + \Delta\mathrm{lat}$	lon_0
4	t_0	$\mathrm{lat}_0 + \Delta\mathrm{lat}$	$\mathrm{lon}_0 + \Delta\mathrm{lon}$
5	$t_0 + \Delta t$	lat_0	lon_0
6	$t_0 + \Delta t$	lat_0	$\mathrm{lon}_0 + \Delta\mathrm{lon}$
7	$t_0 + \Delta t$	$\mathrm{lat}_0 + \Delta\mathrm{lat}$	lon_0
8	$t_0 + \Delta t$	$\mathrm{lat}_0 + \Delta\mathrm{lat}$	$\mathrm{lon}_0 + \Delta\mathrm{lon}$

Given a variable of interest (VOI), (e.g., significant wave height, measured pitch

variance) and waypoint $(\overline{t}, \overline{\text{lat}}, \overline{\text{lon}})$, linear interpolation in three dimensions can be viewed as a weighted summation of the VOI values at each corner. The weight, or scale factor, SF_i, assigned to corner i is defined in Equation 6.2.

$$SF_i = \left(1 - \left|\frac{t_i - \overline{t}}{\Delta t}\right|\right)\left(1 - \left|\frac{\text{lat}_i - \overline{\text{lat}}}{\Delta \text{lat}}\right|\right)\left(1 - \left|\frac{\text{lon}_i - \overline{\text{lon}}}{\Delta \text{lon}}\right|\right) \tag{6.2}$$

Given that the waypoint falls within the bounds of the rectangular prism, Equation 6.2 meets the logical requirement outlined in Equation 6.3, which abides by the principle of energy conservation when interpolating motion or wave elevation variances.

$$\sum_{i=1}^{8} SF_i = 1 \tag{6.3}$$

As shown in Equation 6.2, SF_i is greatest (equal to 1.0) when $(\overline{t}, \overline{\text{lat}}, \overline{\text{lon}})$ falls perfectly on corner i, in which case the remaining seven corners have scale factors equal to zero. As mentioned in Section 6.1.1, the handling of masked wave data corners in this study contrasted that of Chapter V. In this study, when one or more of the eight wave data corners in the dataset were masked for a given 30-minute window, rather than fully omit the sample, SF_i values for the masked corners were set to zero. In this case to ensure that the summation presented in Equation 6.3 held true, the scale factors remaining for the unmasked points were scaled, as shown in Equation 6.4.

$$SF_i = \frac{SF_i}{\sum_{i=1}^{8} SF_i} \tag{6.4}$$

Note that logically, the number of masked corners was always an even number because these points corresponded to latitude and longitude combinations that fell on or too close to land. Given the scale factors for all eight corners, the interpolated VOI could

be calculated using Equation 6.5.

$$\mathrm{VOI}(\bar{t}, \overline{\mathrm{lat}}, \overline{\mathrm{lon}}) = \sum_{i=1}^{8} SF_i \times \mathrm{VOI}(t_i, \mathrm{lat}_i, \mathrm{lon}_i) \tag{6.5}$$

The linear interpolation mechanics described in Equations 6.1-6.5, albeit more detailed here, are equivalent to those used in the previous work of *Schirmann et al.* (2020b) and *Schirmann et al.* (2021), as described in Chapter V; however, the interpolated VOIs differ between this work and the previous work. As noted in Chapter V, the wave data interpolation scheme employed in *Schirmann et al.* (2020b) and *Schirmann et al.* (2021) was imperfect and unreliable in certain cases due to inconsistency in CMEMS' identification of wind waves (WW), primary swell (SW1), and secondary swell (SW2) across the eight corners. For example, primary and secondary swell could swap between corners if the balance of wave energy changed accordingly. This possible inconsistency could lead to unreliable interpolated parameters, particularly in the case of characteristic wave directions. Furthermore, for wave data sources with variable numbers of wave partitions, such as NWW3, there is no reliable consistency between wave partitions at different corners, making the previous approach unfeasible.

In the previous interpolation approach, each of the CMEMS wave parameters was interpolated prior to use as input to the data-driven models. To avoid the drawbacks of the previous approach, this work performed the interpolation step after motions were predicted. More explicitly, the data-driven models in this work generated motion variances at each of the eight corners, and the interpolation scheme detailed above was used to combine the response variances within the model structure itself. The formulation presented in Equation 6.5 is updated to clearly show this combination

approach for heave motion variance, σ_3^2, in Equation 6.6.

$$\sigma_3^2(\bar{t}, \overline{\text{lat}}, \overline{\text{lon}}) = \sum_{i=1}^{8} SF_i \times \sigma_3^2(t_i, \text{lat}_i, \text{lon}_i) \tag{6.6}$$

Motion variance was selected rather than standard deviation to more rigorously preserve energy and follow the same linear superposition principles employed for PBMP generation in the initial twin framework. The incorporation of this interpolation scheme into the shared-layer neural network (SLNN) is detailed in the following section.

6.1.4 Shared-Layer Neural Network (SLNN)

The NNs presented in Chapter IV from *Schirmann et al.* (2019a), *Hageman et al.* (2019), and *MARIN* (2019), and those presented in Chapter V from *Schirmann et al.* (2020b) and *Schirmann et al.* (2021), were fully-connected networks (FCNs), meaning every neuron in each layer was connected to all neurons in the layers before and after it, with the exception of instances where learned weights were equal to zero. As mentioned previously, this conventional NN structure did not enforce similar physics in handling the different directional wave partitions (e.g., WW, $SW1$, and $SW2$ for CMEMS). The model structure developed in this study sought to enforce physical similarity between different wave partitions while allowing variation in the total number of wave partitions between data points. The enforced similarity between wave partitions was achieved through a combination of traditional, linear superposition principles and an input data structure that allowed nonlinear relationships between individual wave partition parameters stored in $\boldsymbol{\ell}_{ij}$ and parameters describing the entire wave system stored in $\mathbf{g}_i$. This new model structure removed the ability of wave partitions to interact directly with one another, which was the cost of allowing variability in the number of wave partitions. Note, however, that the results of this

study for CMEMS show minimal impact on predictive performance from this trade-off, suggesting that the nonlinear relationships allowed between individual partition and system-level wave parameters were sufficient.

The shared-layer NN (SLNN), as its name suggests, shares fully-connected layers between wave partitions. These fully-connected layers can be seen as their own fully-connected network (FCN) within the larger model. First, consider the model structure for a single corner, corner i, used in the interpolation scheme discussed in Section 6.1.3. The wave model data at corner i include global wave parameters stored in $\mathbf{g}_i$ describing the system as a whole and local wave parameters stored in $\boldsymbol{\ell}_{ij}$ describing N_i different wave partitions.

The wave partition-specific variables in $\boldsymbol{\ell}_{ij}$ were joined with the heave, pitch, and roll variance PBMPs in vector $\mathbf{L}_{ij}$ as shown in Equation 6.7. As discussed in Section 6.1.2, these local PBMPs were generated using the parameters specific to wave partition j stored in $\boldsymbol{\ell}_{ij}$.

$$\mathbf{L}_{ij} = \begin{Bmatrix} \boldsymbol{\ell}_{ij} \\ \sigma^2_{3P,ij} \\ \sigma^2_{5P,ij} \\ \sigma^2_{4P,ij} \end{Bmatrix} \begin{matrix} \text{Vector of local, wave partition } j \text{ parameters} \\ \text{Heave PBMP (variance) due to wave partition } j \\ \text{Pitch PBMP (variance) due to wave partition } j \\ \text{Roll PBMP (variance) due to wave partition } j \end{matrix} \tag{6.7}$$

Similarly, the global wave parameters stored in $\mathbf{g_i}$ were joined with the heave, pitch, and roll PBMPs of variance due to all wave partitions in vector $\mathbf{G}_i$, as shown in

Equation 6.8. The vessel's speed, U, was also included in $\mathbf{G}_i$.

$$\mathbf{G}_i = \begin{Bmatrix} \text{U} \\ \mathbf{g}_i \\ \sigma^2_{3P,iT} \\ \sigma^2_{5P,iT} \\ \sigma^2_{4P,iT} \end{Bmatrix} \begin{matrix} \text{Vessel speed} \\ \text{Vector of global wave parameters} \\ \text{Heave PBMP (variance) due to all wave partitions} \\ \text{Pitch PBMP (variance) due to all wave partitions} \\ \text{Roll PBMP (variance) due to all wave partitions} \end{matrix} \tag{6.8}$$

Recall that the heave, pitch, and roll PBMPs stored in $\mathbf{G}_i$ were calculated using linear summation of variances due to the individual wave partitions (stored in $\boldsymbol{\ell}_{ij}$). This application of linear superposition is reiterated in Equation 6.9 for heave.

$$\sigma^2_{3P,iT} = \sum_{j=1}^{N_i} \sigma^2_{3P,ij} \tag{6.9}$$

As mentioned previously, the shared-layer approach leveraged this concept of linear superposition while allowing nonlinear interaction between individual wave partition and global wave system parameters in the model's embedded FCN. For heave, pitch, and roll prediction at corner i, this boiled down to the formulation presented in Equation 6.10, which allowed nonlinear interaction between $\mathbf{G}_i$ and $\mathbf{L}_{ij}$.

$$\boldsymbol{\sigma}^2_{SL,i} = \begin{Bmatrix} \sigma^2_{3SL,i} \\ \sigma^2_{5SL,i} \\ \sigma^2_{4SL,i} \end{Bmatrix} = \sum_{j=1}^{N_i} \mathbf{FCN}\left(\begin{Bmatrix} \mathbf{G}_i \\ \mathbf{L}_{ij} \end{Bmatrix} \right) + \begin{Bmatrix} w_{3P} \\ w_{5P} \\ w_{4P} \end{Bmatrix} \odot \begin{Bmatrix} \sigma^2_{3P,ij} \\ \sigma^2_{5P,ij} \\ \sigma^2_{4P,ij} \end{Bmatrix} \tag{6.10}$$

In Equation 6.10, the FCN is shared between all N_i wave partitions, and the outputs, comprised of a heave, pitch, and roll variance for each partition, are combined through linear superposition. This sharing of layers enforces physics-sharing between wave partitions. The specific FCN structures considered in this study are detailed in Section 6.1.6. Also shown in Equation 6.10, in addition to the FCN, there exists another

term within the summation for corner i. This term is comprised of three weights, each corresponding to a DOF, that are multiplied element-wise (as indicated by the $\odot$ symbol) by the vector of heave, pitch, and roll variance PBMPs for wave partition j. The purpose of this additional term is described in greater detail as it pertains to SLNN initialization in Section 6.1.5. For now, this term can be thought of as a "direct route" for the PBMPs through the model. More explicitly, if these three direct route weight parameters in Equation 6.10 were set to one, as shown for heave in Equation 6.11, and the weights within the FCN were set such that the FCN's output was zero, the output of the SLNN would be exactly equal to the total response variance PBMPs stored in $\mathbf{G}_i$. Rather than travel this direct route themselves, note that the total variance PBMPs resulted from the summation of wave partition-specific PBMPs that followed this direct route. Recall that the PBMPs for wave partition j used in this direct route term were also used in the FCN via vector $\mathbf{L}_{ij}$.

$$w_{3P0} = 1 \tag{6.11}$$

Incorporating the SLNN formulation of Equation 6.10 for a single corner into the interpolation scheme of Equation 6.6, the full SLNN model structure is described in Equation 6.12.

$$\boldsymbol{\sigma}^2_{SL} = \begin{Bmatrix} \sigma^2_{3SL} \\ \sigma^2_{5SL} \\ \sigma^2_{4SL} \end{Bmatrix} = \sum_{i=1}^{8} SF_i \left(\sum_{j=1}^{N_i} \mathbf{FCN}\left(\begin{Bmatrix} \mathbf{G}_i \\ \mathbf{L}_{ij} \end{Bmatrix} \right) + \begin{Bmatrix} w_{3P} \\ w_{5P} \\ w_{4P} \end{Bmatrix} \odot \begin{Bmatrix} \sigma^2_{3P,ij} \\ \sigma^2_{5P,ij} \\ \sigma^2_{4P,ij} \end{Bmatrix} \right) \tag{6.12}$$

The SLNN model was constructed, trained, and tested using `Python`'s `Keras` library with the `TensorFlow` backend (*Chollet et al.*, 2015) as described in Section 6.1.6. While the formulation presented in Equation 6.12 is an accurate representation of the SLNN's conceptual structure, a minor modification was necessary for

practical implementation using `Keras`. Rather than building the model to directly allow variation in the number of wave partitions, N_i, for a given sample and corner, this functionality was achieved using scale factor parameters (i.e., SF_i) at the wave partition level, SF_{ij}, as shown in Equation 6.13.

$$SF_{ij} = \begin{cases} SF_i & \text{if wave partition } j \text{ is defined} \\ 0 & \text{otherwise} \end{cases} \tag{6.13}$$

The summation presented in Equation 6.12 was then redefined to operate over N_{max} wave partitions, as shown in Equation 6.14. N_{max} is the maximum number of wave partitions expected for a given wave data source. In this study, the maximum number of wave partitions observed across AR Train, AR Test, and SR Test for each wave data source, as given in Table 6.1, were used.

$$\boldsymbol{\sigma}^2_{SL} = \begin{Bmatrix} \sigma^2_{3SL} \\ \sigma^2_{5SL} \\ \sigma^2_{4SL} \end{Bmatrix} = \sum_{i=1}^{8} \sum_{j=1}^{N_{\text{Max}}} SF_{ij} \left(\mathbf{FCN} \left(\begin{Bmatrix} \mathbf{G}_i \\ \mathbf{L}_{ij} \end{Bmatrix} \right) + \begin{Bmatrix} w_{3P} \\ w_{5P} \\ w_{4P} \end{Bmatrix} \odot \begin{Bmatrix} \sigma^2_{3P,ij} \\ \sigma^2_{5P,ij} \\ \sigma^2_{4P,ij} \end{Bmatrix} \right) \tag{6.14}$$

As given in Table 6.1, for the CMEMS data, N_{max} was equal to three (WW, SW1, and SW2). Although the CMEMS output format was rigid, existence (i.e., non-zero wave energy) of all three partitions was not the case for many data points. While a seemingly logical alternative to this SF_{ij} approach would be to feed the FCN with wave energy and PBMPs equal to zero for a non-existent wave partitions (as was done for previous NNs), assigning placeholder values to other parameters (e.g. period, direction) could negatively impact FCN training.

As mentioned previously, given the manageable number of wave partitions for the CMEMS dataset, a more traditional NN, nearly identical to those of Chapter V, was also constructed and trained for comparison to the SLNN. While this NN was fully-

connected and used the same variables as the NNs of Chapter V, it incorporated the interpolation scheme described for the SLNN. The adaptation of this scheme for the three wave partitions of the CMEMS dataset is demonstrated in Equation 6.15. Although not attempted in this study, for wave data sources with more wave partitions, each additional partition would append one partition-specific vector (e.g., $\mathbf{L}_{ij}$) to the FCN input vector in Equation 6.15.

$$\boldsymbol{\sigma}^2_{NN} = \begin{Bmatrix} \sigma^2_{3NN} \\ \sigma^2_{5NN} \\ \sigma^2_{4NN} \end{Bmatrix} = \sum_{i=1}^{8} SF_i \left(\mathbf{FCN} \left(\begin{Bmatrix} \mathbf{G}_i \\ \mathbf{L}_{iWW} \\ \mathbf{L}_{iSW1} \\ \mathbf{L}_{iSW2} \end{Bmatrix} \right) \right) \tag{6.15}$$

In Equation 6.15, note that the wind sea or swell indicator variable Q was dropped from vectors $\mathbf{L}_{ij}$ because weights were not shared between wave partitions. Equation 6.15 also visually emphasizes how a network with more than these three wave partitions (e.g., 13 for NWW3 or 24 for CDS) could lead to a very large NN that would be difficult or impossible to train for reliable response prediction. For reference, Table 6.5 summarizes the the PBMPs generated using the initial twin framework and the models constructed, trained, and tested for each wave data source in this study.

Table 6.5: Summary of the PBMPs generated using the initial twin framework and the models constructed, trained, and tested for each wave data source.

Wave Data Source	Model
CMEMS	PBMPs
	NN
	SLNN
CDS	PBMPs
	SLNN
NWW3	PBMPs
	SLNN

6.1.5 SLNN Weight Initialization

As mentioned in the previous section, if the three direct route weights in Equation 6.10 were set to values of one and the FCN weights were set such that FCN output was always zero, the three outputs of the SLNN would perfectly match the total heave, pitch, and roll variance PBMPs provided as model input for corner i. Carrying this direct route behavior through Equations 6.12 and 6.14, the output of the SLNN would be interpolated total heave, pitch, and roll variance PBMPs. The advantage of this behavior would be initialization of the `Keras` model at the exact performance level provided by the PBMPs themselves, as detailed in the following paragraph.

As in previous studies, the MSE loss function was employed for model training. As expected, randomly initializing the FCN and direct route weights using a standard method (He Normal, in this case) yielded large MSE values prior to training (*He et al.*, 2015). In general, these MSE values can be viewed as a starting point for the learning process, which seeks to reduce the MSE loss. Given the number of simultaneous weight calculations considered in training a network with shared-layers, the experiences gathered from this study showed that SLNN training can be more time intensive than traditional NN training and very susceptible to entrapment in local minima with MSE values far greater than those of the PBMPs themselves. As such, the goal of this combined direct route and initialization approach was to start model training with the MSE loss function at exactly the value provided by the PBMPs themselves. The purpose was to encourage faster and more stable improvement upon the PBMPs rather than spending training time in an effort just to reach the quality-level of the PBMPs.

The initialization method employed in this study relied upon the output of the FCN being approximately zero for all input variables to start. However, in order for a network to learn, it is extremely important that weights are set to non-zero values that vary between neurons. Therefore, setting all network weights to zero was not

an option. Furthermore, it is well-known in the machine learning community that near-zero weights in a network layer can prevent network backpropagation algorithms from updating weights in previous (i.e., earlier in the network) layers. As such, it would be unwise and inappropriate to set the final FCN layer weights to near-zero values because it could prevent the network from learning the other network weights. Therefore, the only neurons that were set to near- (but still non-zero) values were the weights between the input layer and first hidden layer (HL). To achieve this behavior, the standard weights assigned by the He Normal initializer were divided by 1000. This denominator was selected through experimentation such that the SLNN's initial MSE was nearly identical to that of the PBMPs themselves but the network could still learn effectively.

After developing and experimenting with this initialization strategy for SLNN model-training, another potential improvement came to light. Given their poor quality, the roll PBMPs tended to dominate the initial MSE value, which was the mean of heave, pitch, and roll variance PBMPs. As a result, at the beginning of training, the heave and pitch MSE values would often climb well above their initial PBMP-based values despite the initialization as the network improved upon the roll MSE. To combat this behavior, initial direct route weights were selected that were not necessarily equal to one. More specifically, in contrast to Equation 6.11, the approach initialized the heave, pitch, and roll PBMP direct route weights using Equation 6.16, which demonstrates the calculation for heave.

$$w_{3P0} = \frac{\overline{\sigma_{3M}^2}}{\left(\overline{\sum_{i=1}^{8} SF_i \times \sigma_{3P,iT}^2}\right)} \tag{6.16}$$

Equation 6.16 shows that the direct route weights were initialized to the mean measured variance divided by the mean interpolated total variance PBMP across all AR Train samples for a given DOF. These weights can be thought of as a slope by which

the PBMPs were multiplied to yield a new predicted value. If the mean predicted and measured variances were equivalent, this weight would equal one. As such, heave and pitch yielded initial weight values much closer to one than the low-quality roll PBMPs. This interpolation scheme lowered the initial roll MSE value for each wave data source well below that of the underlying PBMPs, making it more similar to the heave and pitch MSE values. As a result, there was less detriment to the heave and pitch DOFs in the early stages of model training. The direct route weights pre- and post-training are presented in Section 6.2. Furthermore, the MSE values for each DOF are compared between the PBMPs and the pre-trained SLNN to demonstrate this initialization approach.

6.1.6 Model Construction, Training, and Cross-Validation (CV)

As mentioned previously, SLNN and NN construction, training, and testing was performed using `Python`'s `Keras` library with the `TensorFlow` backend (*Chollet et al.*, 2015). `Keras`' functional API made implementation of shared FCN layers and direct route weights (for the SLNN, specifically) possible within a single model. Although the NN did not share FCN weights between wave partitions, the weights were still shared between interpolation corners as shown in Equation 6.15, and therefore also required `Keras`' shared-layer capabilities.

For the CMEMS SLNN, the FCN had seven input variables corresponding to the four local wave parameters listed in Table 6.3 and the three wave partition-specific PBMPs, as shown in Equation 6.7. For the CDS SLNN, the FCN had six input variables corresponding to the local input variables and local PBMPs, one less than CMEMS due to the lack of a wave partition type indicator. For both CMEMS and CDS, an additional ten FCN input variables resulted from the global wave parameters listed in Table 6.2, the associated total PBMPs, and the vessel's speed, as shown in Equation 6.8. Therefore, there were 17 total input variables to

the FCN of the CMEMS SLNN and 16 total input variables to the FCN of the CDS SLNN. Also shown in Tables 6.2 and 6.3 and consistent with Equations 6.7 and 6.8, for NWW3, this amounted to 19 total input variables (nine local and ten global). The 17 total FCN input variables for the CMEMS SLNN contrasted the 28 input variables necessary for the NN in this study and those of Chapter V. This study followed the same input and output variable normalization procedure (adapted for variance instead of amplitude) outlined in Section 5.1.7. Note that wave parameter input variable normalization was wave data source-dependent, but by definition, PBMP and output variance normalization was consistent across all three wave data sources.

The NN and SLNN FCNs were composed of multiple hidden layers with varying numbers of neurons, and an output layer with three neurons corresponding to heave, pitch, and roll variances. As in Chapter V, training was performed using the MSE loss function, Adam optimizer, and a batch size of 50. Both the input and output layer neurons used linear activation functions. The FCN HLs used neurons with the leaky rectified linear unit (Leaky ReLU) activation function. As in Chapter V, $L2$ weight regularization was applied between layers of the FCN to mitigate overfittting. Note that the regularization penalty was not applied to the direct route weights of the SLNNs. The FCN regularization parameter, λ, and HL architectures for the NN and SLNNs were determined via cross-validation as described in the following paragraph.

For the three SLNNs and single NN constructed in this work, 10-fold CV was employed to select λ and HL architectures for the FCNs. Recall that Chapter V used 5-fold CV; while the procedures of 5- and 10-fold CV (described in 5.1.10) match, more folds (i.e., subsets) are typically preferred because increasing the number of unique sample sets is beneficial from a statistical viewpoint. Of course, it is necessary to keep the number of folds within reason such that the subsets are of sufficient size and the CV process is computationally tractable. Given the 10,034 samples of AR Train, 10-fold CV amounted to approximately 1,003 samples per subset in this study.

Note that due to the input data structure necessary for these shared-layer models, use of `scikit-learn`'s `GridSearchCV` was not possible as implemented in Chapter V (*Pedregosa et al.*, 2011). Therefore, a custom function was written with similar capabilities to `GridSearchCV` for NN and SLNN model selection. Once again, the CV procedures were performed using computational resources and services provided by Advanced Research Computing at the University of Michigan, Ann Arbor.

As detailed in Section 5.1.10, although random partitioning is considered best practice before partitioning into CV subsets for many datasets, through experimentation, it was determined that similarities between sequential points in the train dataset yielded overfitting despite the regularization parameter. As promised in Section 5.1.10, this topic was reconsidered in this study. Rather than rely on fixed subsets of the chronologically partitioned dataset as done in Chapter V, this study introduced a hybrid CV partitioning approach, which is described in the following paragraph.

The chronologically-ordered dataset was partitioned into 10 subsets of approximately equal size, and the standard CV procedure described in Section 5.1.10 was followed for all combinations of candidate HLs and values of λ. In contrast to Chapter V, rather than ending the CV procedure here and selecting the model with the lowest average CV loss across the 10 subsets, the entire procedure was repeated several more times; however, for each repetition, the starting point for subset partitioning was shifted cyclically such that the subsets changed (i.e., the partitions between subsets fell in different locations) without shuffling. More specifically, given a total of L or $L+1$ (1,003 or 1,004 for AR Train) training samples in each of the 10 CV subsets and R total repetitions, the cyclical shift, s, for repetition r was defined as shown in Equation 6.17.

$$s = \text{floor}\left(\frac{(r-1)L}{R}\right) \tag{6.17}$$

The approach outlined in Equation 6.17 retained that similar, sequential points stayed together (unless they fell on a partition boundary) to avoid the aforementioned pitfalls

of random partitioning for this application, but still considered a variety of subsets, which was a distinct advantage over the previous chronological approach. The CV losses for each combination of λ and HL architecture were then averaged across all 10 subsets and R repetitions for selection of the best candidate. Due to the necessary lack of randomness in the developed CV approach to maintain chronology, the number of repetitions was held fixed at six for all candidates of a given model. This number was primarily motivated by computing time with regard to the number of candidates desired for each run and the number of CPUs employed for parallel processing. The model with the lowest average CV loss was then selected and retrained using all data in AR Train. Note that this retraining process was performed 10 times, and the retrained model with the lowest training loss was selected as the final model for a given wave data source. The results of this CV selection process for each wave data source and model are presented in the following section.

6.2 Results

Table 6.6 summarizes the best model(s) for each wave data source determined via the employed 10-fold CV approach. The average CV loss, included in the first column of Table 6.6, was the performance metric used to determine the best combination of λ and HL architecture. The standard deviation of CV loss column shows that there was notable variation in model performance across different subsets and CV repetitions for all models. The average CV loss and standard deviations of CV loss were based on the normalized output motion variances used by the models, which contrasts the use of motion amplitudes in previous chapters. As in Chapter VI, due to the computational cost of training neural networks, the number of user-defined candidates considered for selection of the best models was limited, and more optimal models may exist. Nevertheless, the average CV losses resulting from significantly different HL structures were often similar to these best model values, meaning model performance on train

data was not particularly sensitive to the HL architecture selected. In Table 6.6, it is worth noting that the average CV losses were coincidentally ordered from least to greatest. Although CV losses may be indicative of model performance on train data, model performance conclusions were saved for discussion of the AR Test and SR Test results.

Table 6.6: Results of the 10-fold CV model selection processes for each data-driven model. These selections were performed exclusively using the data in AR Train.

Wave Data Source	Model	Average CV Loss	Std. Dev. of CV Loss	λ	HL Architecture
CMEMS	NN	0.001666	0.001294	0.000010	[120, 120]
	SLNN	0.001675	0.001176	0.000130	[100, 100]
CDS	SLNN	0.001905	0.001302	0.000530	[60, 60]
NWW3	SLNN	0.001942	0.001097	0.000055	[100, 100, 100]

Table 6.7 shows the heave, pitch, roll, and total variance MSE values for AR Train using the three different wave data sources. In addition to the MSE values for the PBMPs and the data-driven model(s), each data source includes the MSE values generated by the direct route slopes prior to model training (SL Slope rows) when model outputs were based solely on the PBMPs multiplied by the initial direct route weights (defined in Equation 6.11). Reviewing these rows of Table 6.7, it is clear that this initialization approach was slightly detrimental for the heave and pitch DOFs in several instances relative to the PBMPs; however, the roll variance MSE values were drastically reduced for all three wave data sources, which significantly decreased the total variance MSE prior to model training and was the desired outcome of the developed initialization approach. The direct route slopes of the SLNNs pre- and post-training are presented in Table 6.8. Although conclusions that can be drawn from these slopes are limited given that the FCN could also account for linear relationships between PBMPs and model output, it is interesting to see the relative weights applied to each DOF's PBMPs on this direct path. As described in Section 6.1.5, this direct

route was critical for model initialization such that the SLNNs consistently improved upon the underlying PBMPs. Note that the SL Slope rows of Table 6.7 were included primarily for demonstration of the initialization approach rather than as a viable independent model; therefore, SL Slope results are not presented for the AR Test or SR Test datasets in this discussion.

Table 6.7: AR Train performance of the data-driven models and their underlying uncorrected PBMPs. As indicated, the MSE values are those of the normalized motion variances, which were used in model training, rather than normalized motion amplitudes. The SL Slope entries correspond to the model output before training based solely on the initial direct route slopes defined in Equation 6.11.

Wave Data Source	Model	Heave Var. MSE [-]	Pitch Var. MSE [-]	Roll Var. MSE [-]	Total Var. MSE [-]
CMEMS	PBMPs	0.002541	0.002072	0.069400	0.024671
	SL Slope	0.003040	0.004123	0.002807	0.003328
	NN	0.000602	0.000945	0.000579	0.000709
	SLNN	0.001089	0.001713	0.001184	0.001330
CDS	PBMPs	0.002263	0.002190	0.117271	0.040575
	SL Slope	0.001267	0.002240	0.003683	0.002400
	SLNN	0.001235	0.002099	0.001403	0.001581
NWW3	PBMPs	0.002779	0.002594	0.089732	0.031702
	SL Slope	0.003043	0.004259	0.004605	0.003974
	SLNN	0.001179	0.001866	0.001136	0.001396

Table 6.8: SLNN direct route slopes (SL Slopes) pre- and post-training.

.

Wave Data Source	Model	**Heave** Pre-	**Heave** Post-	**Pitch** Pre-	**Pitch** Post-	**Roll** Pre-	**Roll** Post-
CMEMS	SLNN	1.577	1.016	1.061	0.680	0.234	0.105
CDS	SLNN	1.460	1.388	0.966	0.880	0.169	0.049
NWW3	SLNN	1.650	1.060	0.996	0.602	0.215	0.037

Although insights are limited when analyzing results for a train dataset, it is clear from Table 6.7 that the NN and SLNN models learned to improve upon the underlying PBMPs of the AR Train dataset in each DOF with respect to the PBMPs. Nevertheless, analyses of test data performances in the following discussion are far

more indicative of real-world model performance.

Table 6.9 shows the heave, pitch, roll, and total variance MSE values for AR Test using the three different wave data sources. For all three DOFs and each wave data source, the data-driven model(s) outperformed their respective PBMPs, showing the models' efficacy for improved response prediction. Recall that significant differences between wave data source resolutions, formats, and numbers of considered windows made direct comparisons across the wave data sources inappropriate; however, the results for these three diverse wave data sources do demonstrate the versatility of the developed SLNN structure and the robustness of this approach.

For the CMEMS results in Table 6.9, the SLNN yielded the lowest MSE for heave, and the NN yielded the lowest MSE for pitch and roll. For all three DOFS, the NN and SLNN MSE values were more similar to each other than to the PBMPs that they improved upon. Nevertheless, the NN did yield a lower total MSE for the AR Test dataset. Given the required removal of direct interaction between wave partitions to enforce physics-sharing in the SLNN structure, it is unsurprising that the NN slightly outperformed the SLNN for the CMEMS dataset. Nevertheless, a traditional NN structure for CDS and NWW3, which involved significantly more wave partitions than CMEMS, would be much more difficult or impossible to train for reliable response prediction. Therefore, the similarity between performances of these two model types shows promise for the SLNN model structure.

Table 6.10 shows the heave, pitch, roll, and total variance MSE values for SR Test using the three different wave data sources. For roll, the data-driven models outperformed the PBMPs for each of the three wave data sources, and the NN once again outperformed the SLNN for CMEMS. In contrast to the AR Test results of Table 6.9, the heave and pitch PBMPs yielded lower MSE values than the data-driven model(s) in several instances. However, it is important to note that the MSE values of the underlying PBMPs for SR Test are far lower than their counterparts for

Table 6.9: AR Test performance of the data-driven models and their underlying PBMPs. As indicated, the MSE values are those of the normalized motion variances, which were used in model training, rather than normalized motion amplitudes. The SL Slope entries correspond to the model output before training based solely on the initial direct route slopes defined in Equation 6.11.

Wave Data Source	Model	Heave Var. MSE [-]	Pitch Var. MSE [-]	Roll Var. MSE [-]	Total Var. MSE [-]
CMEMS	PBMPs	0.003651	0.002268	0.078923	0.028280
	NN	0.001235	0.001351	0.003002	0.001865
	SLNN	0.001160	0.001566	0.003705	0.002146
CDS	PBMPs	0.003860	0.001917	0.082919	0.029565
	SLNN	0.001613	0.001886	0.004886	0.002804
NWW3	PBMPs	0.004136	0.006239	0.139208	0.049861
	SLNN	0.002159	0.003127	0.004074	0.003091

AR Test. Furthermore, the motion variance MSE values considered in these tables do not necessarily tell the full story of model performance, as detailed in the following paragraphs.

Table 6.10: SR Test performance of the data-driven models and their underlying PBMPs. As indicated, the MSE values are those of the normalized motion variances, which were used in model training, rather than normalized motion amplitudes. The SL Slope entries correspond to the model output before training based solely on the initial direct route slopes defined in Equation 6.11.

Wave Data Source	Model	Heave Var. MSE [-]	Pitch Var. MSE [-]	Roll Var. MSE [-]	Total Var. MSE [-]
CMEMS	PBMPs	0.000363	0.000323	0.005781	0.002156
	NN	0.000737	0.000394	0.000477	0.000536
	SLNN	0.000555	0.000357	0.000595	0.000503
CDS	PBMPs	0.000339	0.000232	0.004925	0.001832
	SLNN	0.000448	0.000231	0.000587	0.000422
NWW3	PBMPs	0.000341	0.000458	0.012194	0.004331
	SLNN	0.000566	0.000359	0.000664	0.000530

Figure 6.2 is similar to Figure 5.13 of Chapter V, which showed MSE of predicted heave, pitch, and roll amplitudes as a function of their respective (binned) measured motion amplitudes using normalized values, as well as the underlying histograms for

each dataset. However, rather than MSE, Figure 6.2, shows the root mean square percentage error (RMSPE) of predicted heave, pitch, and roll amplitudes for the CMEMS wave data. This RMSPE calculation is demonstrated in Equation 6.18 for SLNN-predicted heave amplitude, σ_{3SL}.

$$RMSPE = 100 \times \sqrt{\frac{\sum_{k=1}^{K} \left(\frac{\sigma_{3M,k} - \sigma_{3SL,k}}{\sigma_{3M,k}} \right)^2}{K}} \tag{6.18}$$

In Equation 6.18, K is the total number of points in a given motion amplitude bin of Figure 6.2. The reasoning behind RMSPE for this chapter was to show the performance of the data-driven models relative to their underlying amplitudes. As mentioned in Chapter V, the purpose of looking at model performance as a function of measured amplitudes was to quantify predictive performance in more severe operating conditions, when a vessel owner or operator may care most about model reliability. This RMSPE approach takes the analysis one step further and examines model performance in more understandable terms of percentages.

In Figure 6.2, the most striking features at first glance are the large RMSPE values for the PBMPs and both data-driven models in the lowest motion amplitude bins, where the curves extend above the plots' 100% RMSPE threshold in several instances. There are two main reasons behind this behavior. First, in these lower heave, pitch, and roll amplitude bins, the measured values are small, by definition. As such, the RMSPE values can be very large even with relatively small errors between measured and predicted values because the denominator of Equation 6.18 is small. Second, for the NN and SLNN models, specifically, these models were trained using motion variance MSE rather than motion amplitude MSE. In general, MSE puts more weight on larger errors than smaller ones, which contrasts some alternative metrics such as mean absolute error. This behavior of MSE is typically desirable for regression applications, such as this one. That said, the use of motion variances instead

of amplitudes for model training put even greater weight on larger amplitude errors than seen for the models of Chapter V. Despite the relatively large number of data points in these lower amplitude bins (shown by the underlying histograms of Figure 6.2), the models still placed a lot of weight on accuracy at higher amplitudes during the training process, as shown by the AR Train plots. Furthermore, given the importance of reliable response predictions for more severe motion amplitudes mentioned above, poorer model performances in the lower amplitude bins may not be overly concerning. To put the prediction errors for lower motion amplitude bins in context, Figure 6.3 shows the more traditional root mean square error (RMSE) metric for the predicted normalized motion amplitudes using CMEMS wave data. Figure 6.3 shows the same general relationships between model and PBMP performances in greater amplitude bins as Figure 6.2's RMSPE plots, but also that the error magnitudes in lower amplitude bins are not particularly dramatic.

Excluding the lowest amplitude bins, the RMSPE plots of Figure 6.2 show that the SLNN and NN models both outperformed the underlying PBMPs in most instances for both AR Test and SR Test. These improvements were most pronounced for roll, which had very poor underlying PBMPS. However, there were still notable improvements for heave and pitch. For both heave and pitch in the AR Test and SR Test datasets, the performances of the NN and SLNN were very similar in most bins, with a possible slight advantage shown for the NN in select instances, consistent with the discussion of Table 6.9. For SR Test in particular, the strong heave and pitch performances imply that the motion variance MSE relationships depicted in Table 6.10 were misleading, which was likely due to the relatively large amount of SR Test data points at low amplitudes and the variance MSE metric's definition. For AR Test and SR Test roll, the NN's advantage was slightly more pronounced, especially at higher amplitudes, but the SLNN still dramatically reduced the RMSPE of the underlying PBMPs. Given the nonlinear complexities of roll in comparison to heave

and pitch, the better performance for the NN is unsurprising based on the SLNN's purposefully more restrictive structure. As discussed previously, the NN's slight advantage over the SLNN in all DOFs is also unsurprising given that the CMEMS wave data only provided a maximum of three wave partitions. However, the feasibility of a well-trained NN declines as the maximum number of wave partitions increases, which is exactly why NNs were not attempted for CDS or NWW3. The most important takeaway from the relative relationships presented in Figures 6.2 and 6.3, as well as Tables 6.9 and 6.10, is that the SLNN is a worthy competitor for the NN and significantly improves upon the underlying PBMPs using its physics-informed structure. Furthermore, this conclusion was valid for both the *Neil Armstrong* and *Sally Ride*, showing the potential application of these models for fleet-wide response prediction.

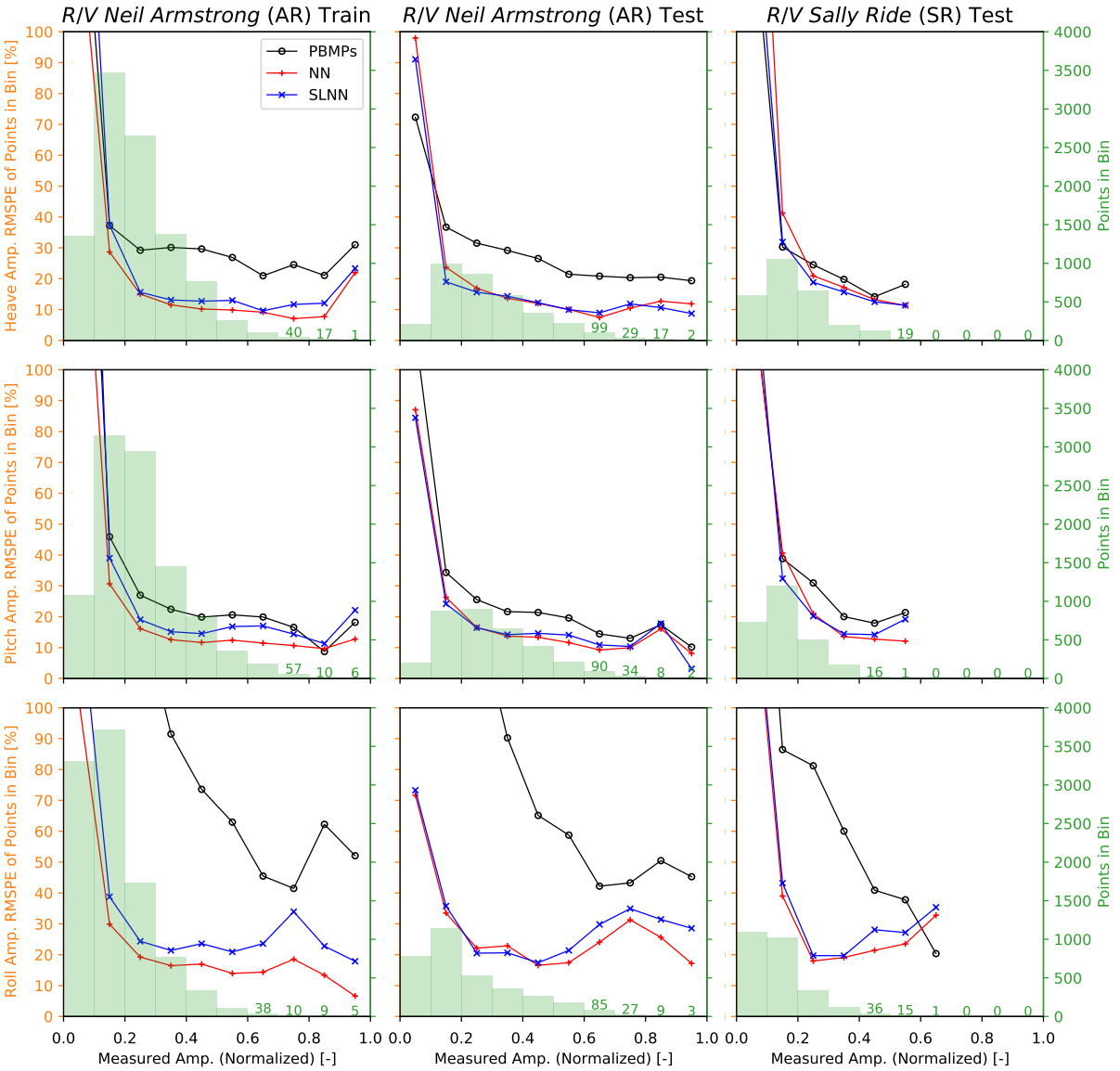

Figure 6.2: Plots of heave, pitch, and roll RMSPE as a function of their respective normalized motion amplitude measurements for the CMEMS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSPE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

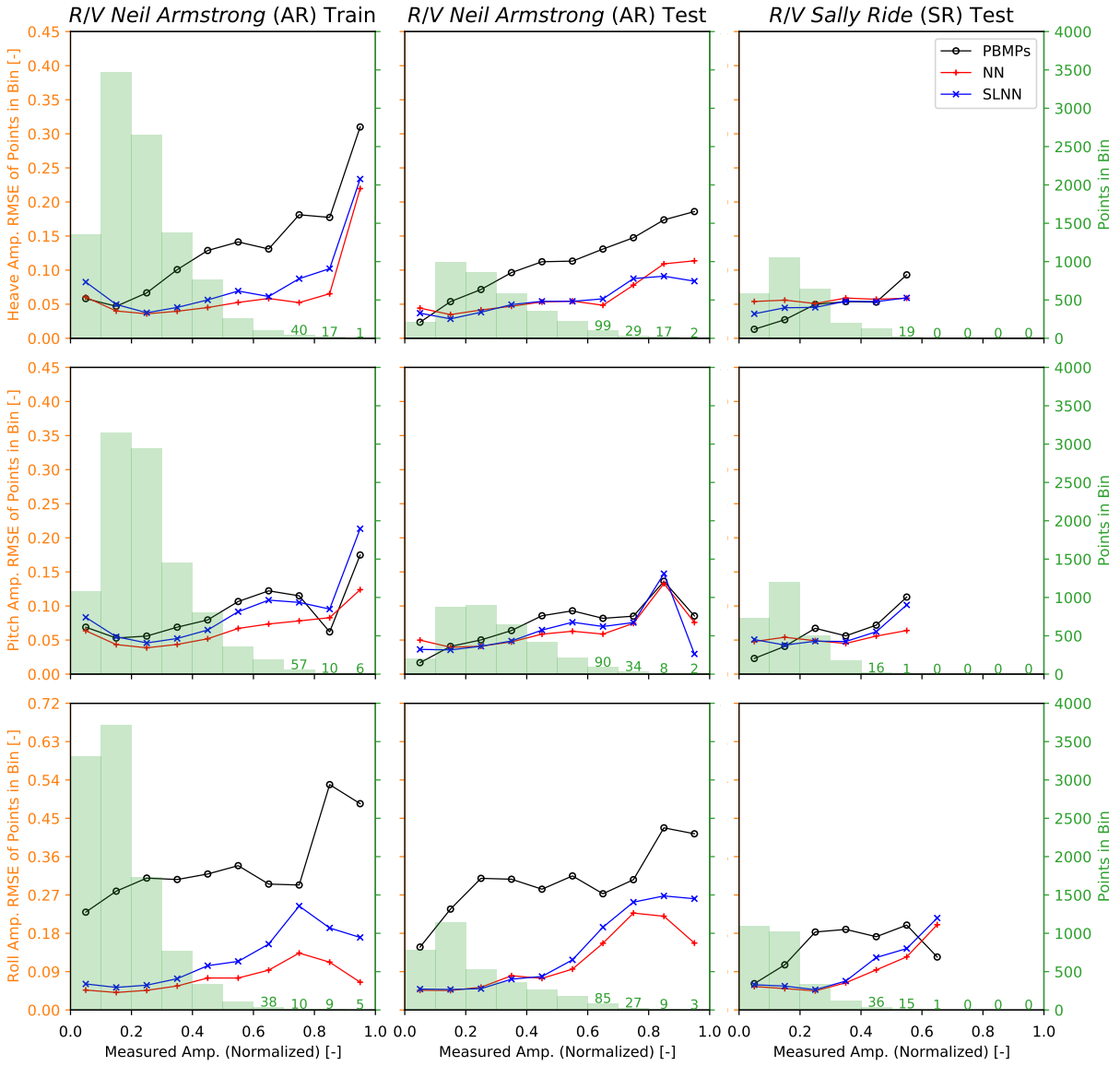

Figure 6.3: Plots of heave, pitch, and roll RMSE as a function of their respective normalized motion amplitude measurements for the CMEMS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

Figures 6.4 and 6.5 show the RMSPE and RMSE results for the SLNN and underlying PBMPs using CDS wave data. Comparing the lowest motion amplitude bins in these two figures, it is evident that small measured motions again yielded large RMSPE values despite RMSE values that were not outlandishly large. Beyond these low amplitude bins, the SLNN typically outperformed the underlying PBMPs for both AR Test and SR Test, with significant performance improvement seen for the heave and roll DOFs. The less significant improvement seen for pitch was attributed to the relatively high quality of the pitch PBMPs. More specifically, the CDS 2D wave spectra removed the need for idealized wave spectra construction from characteristic parameters. Given the high quality of the pitch PBMPs discussed in Chapter V, which implied that the pitch RAOs were relatively accurate, it is not surprising that the addition of defined spectral shapes for 24 wave directions yielded solid PBMPs. Still, the SLNN managed to improve upon these PBMPs in several instances. The most prominent exceptions to the SLNN's superior performance happened for a few roll bins where a small number of train and test windows were considered, as portrayed by the histograms. AR Train contained similar behavior to AR Test in these roll bins with few train data points, showing consistency in the trained model. Therefore, it is likely that these blips in performance were due to train data sparsity. Nevertheless, for the majority of windows, the SLNN significantly reduced roll RMSPE and RMSE relative to the underlying PMBPs. The success of the SLNN in all three DOFs using this unique wave data source with 24 wave partitions, where a traditional NN structure would be far less feasible for reliable response prediction, demonstrates the developed model's versatility and robustness.

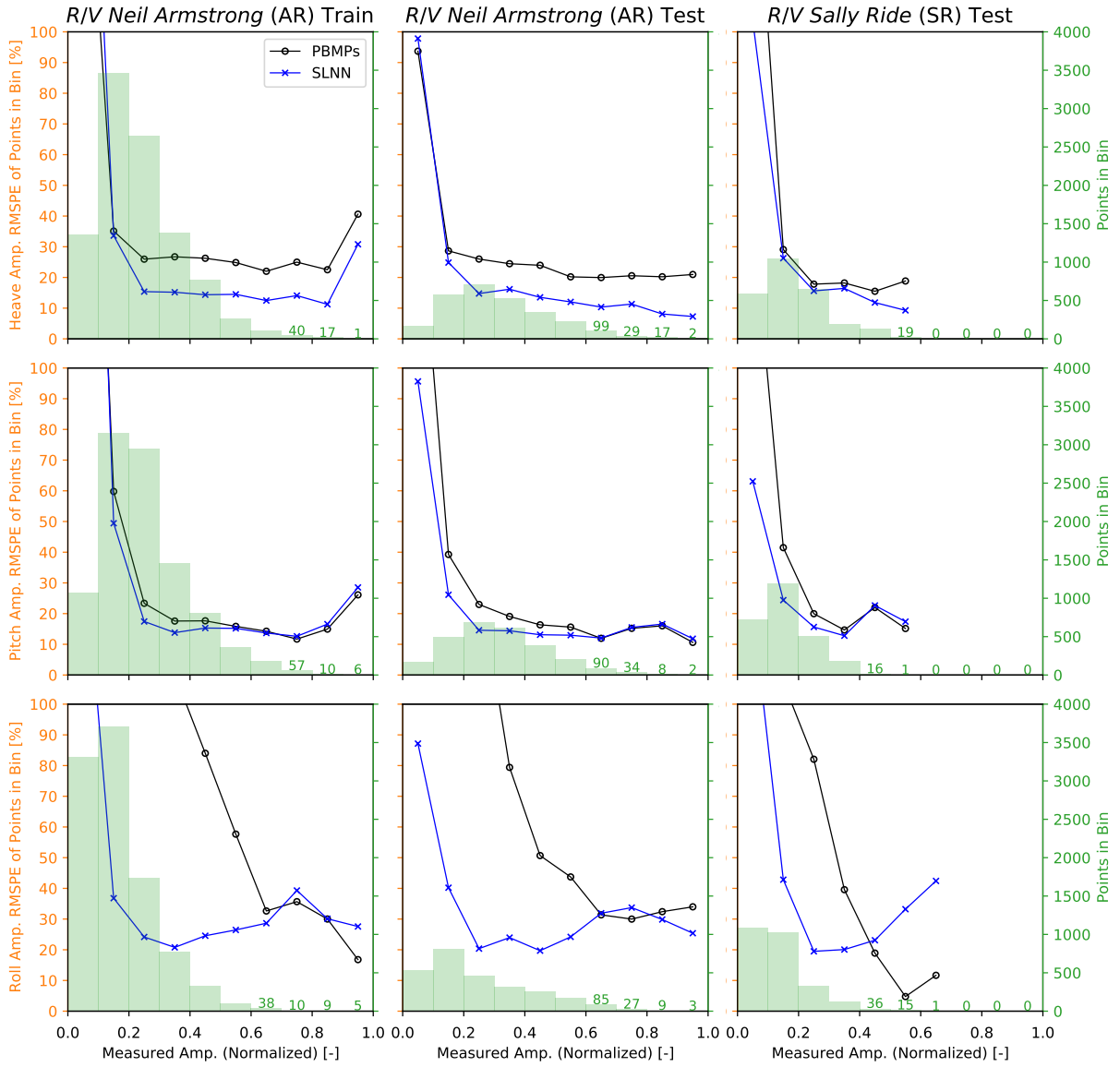

Figure 6.4: Plots of heave, pitch, and roll RMSPE as a function of their respective normalized motion amplitude measurements for the CDS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSPE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

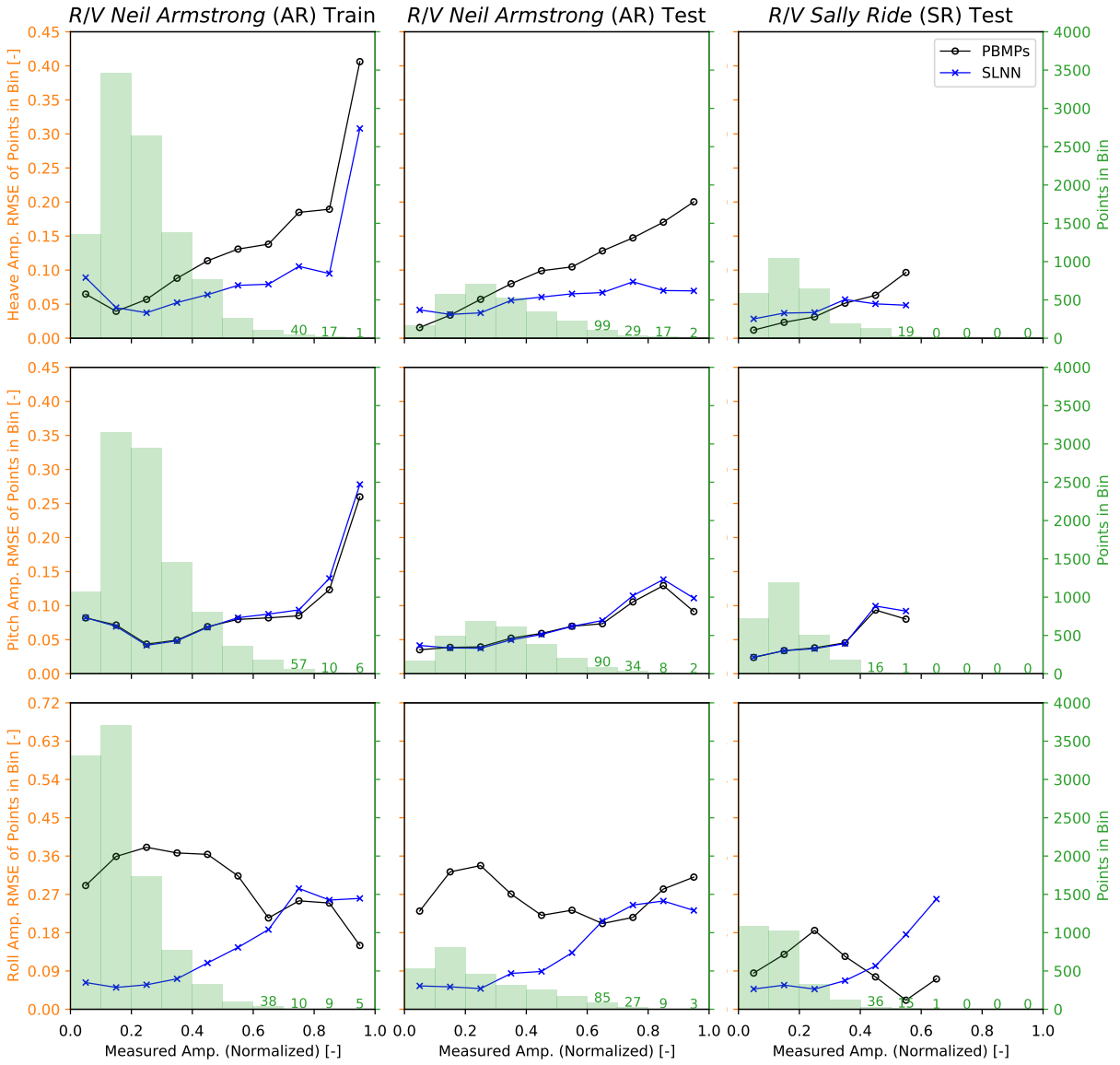

Figure 6.5: Plots of heave, pitch, and roll RMSE as a function of their respective normalized motion amplitude measurements for the CDS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

Figures 6.6 and 6.7 show the RMSPE and RMSE results for the SLNN using NWW3 wave data. Consistent with the results for CMEMS and CDS, comparing the lowest motion amplitude bins' RMSPE and RMSE behavior, large RMSPE values were due primarily to small measured motion amplitudes. Beyond these low amplitude bins, Figures 6.6 and 6.7 show that the SLNN notably outperformed the underlying PBMPs with very few exceptions, and these exceptions all occurred in bins with a very small number of AR Train data and very few AR Test or SR Test samples.

Although CDS had up to 24 wave partitions, in order for a given direction to be worthy of its own wave partition, only one wave frequency bin with non-zero energy was necessary. As such, 24 partitions were frequently considered. Therefore, of the three wave data sources, NWW3 had the largest variability in number of wave partitions, which extended up to 13 observed in this study. Given this large amount of variability in the number of NWW3 wave partitions, the SLNN's impressive performance using NWW3 wave data puts great emphasis on the developed structure's versatility and robustness where a traditional NN would be far less feasible.

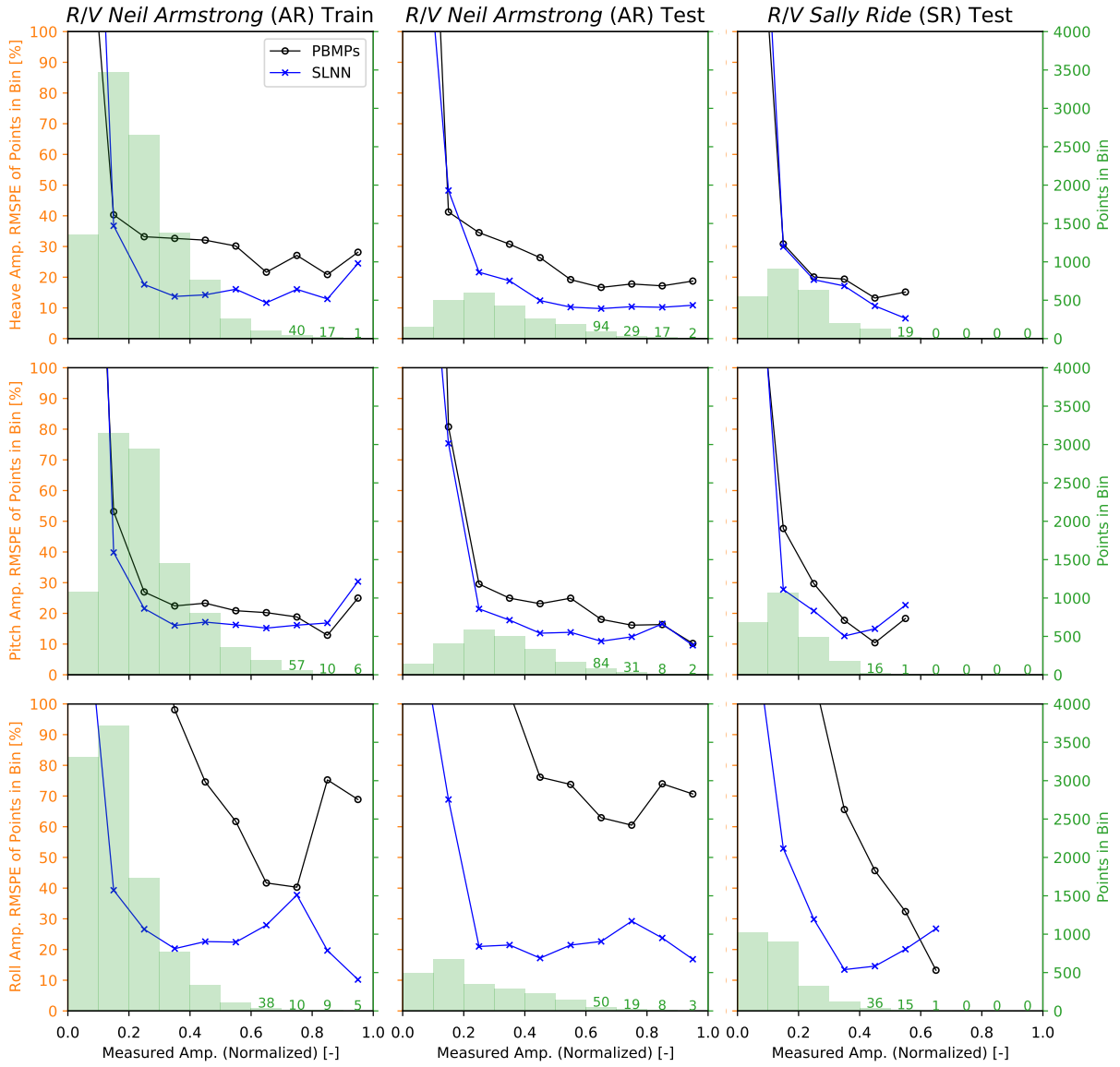

Figure 6.6: Plots of heave, pitch, and roll RMSPE as a function of their respective normalized motion amplitude measurements for the NWW3 wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSPE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

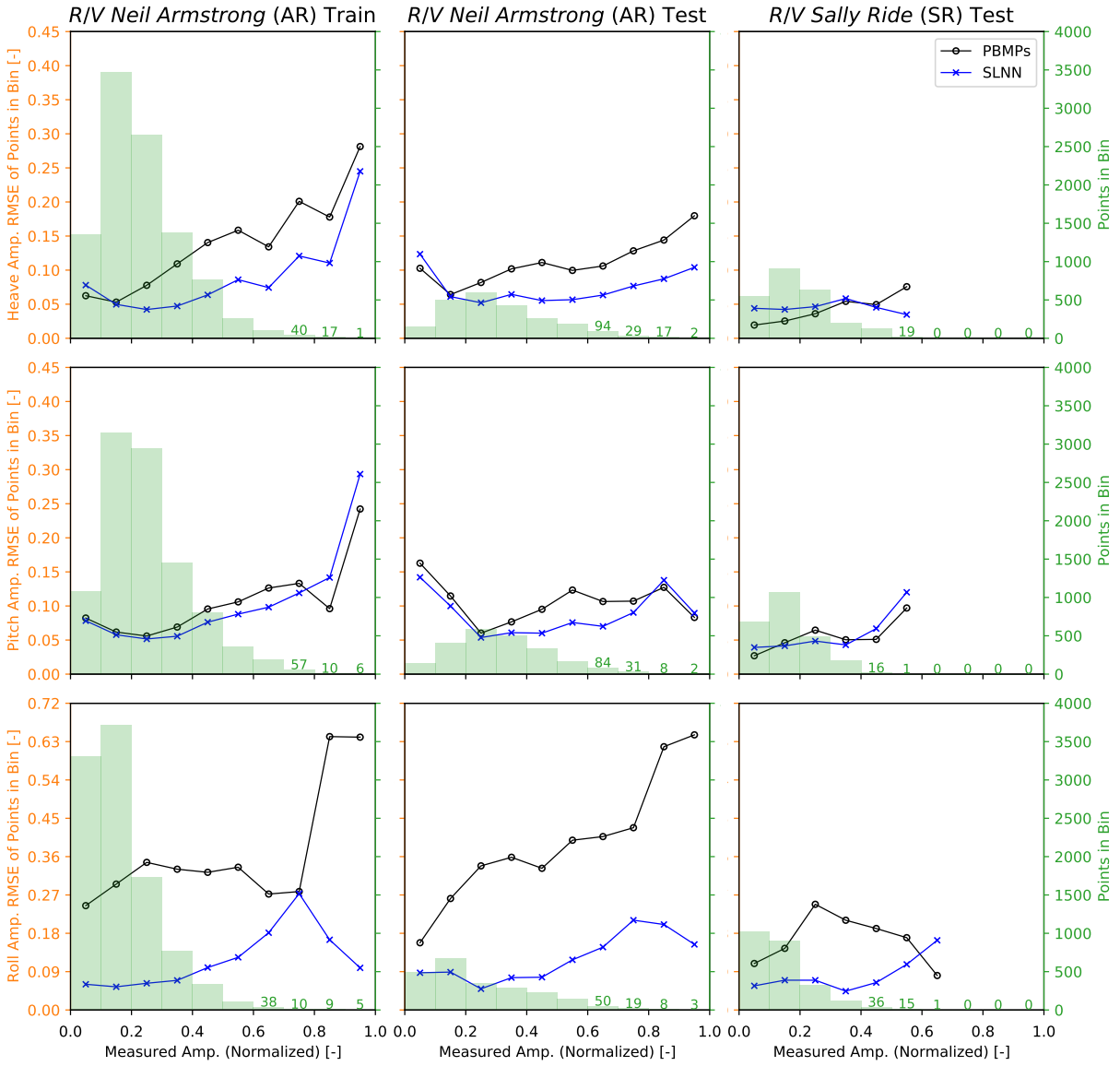

Figure 6.7: Plots of heave, pitch, and roll RMSE as a function of their respective normalized motion amplitude measurements for the NWW3 wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

6.3 Shared-Layer NN Conclusions

This chapter presented a novel model structure that shared physics between different directional wave partitions. In addition to enforcing consistent physics, the developed SLNN structure allowed variation in the number of wave partitions between samples, which was demonstrated for three unique wave data sources with maximum numbers of wave partitions ranging from 3 to 24. This model structure was also advantageous because the number of input variables was not dependent on the number of wave partitions, which contrasts the more traditional NN structure of Chapter V. Finally, the developed shared-layer structure incorporated a robust interpolation approach that avoided the interpolation shortcomings of Chapter V. Combination of this improved interpolation approach and the shared network layers was paramount for effective implementation of SLNN models using the CDS and NWW3 datasets, which had maxima of 24 and 13 wave partitions observed in this study, respectively.

Through analysis of heave, pitch, and roll RMSPE and RMSE values as a function of their respective measured amplitudes for each wave data source, it was shown that the SLNN outperformed the underlying PBMPs with few exceptions. Although there were instances for CMEMS where the NN outperformed the SLNN, especially for roll, this behavior was not surprising given the purposefully restrictive structure of the SLNN model. The fact that the SLNN competed with the NN and significantly improved upon the underlying PBMPs in most cases shows great promise. This promise was further highlighted by the SLNN's success for the CDS and NWW3 wave datasets, where a traditional NN structure would likely struggle due to the large and variable number of wave partitions. The SLNN's impressive performance was seen for both the *Neil Armstrong* and *Sally Ride* test datasets, which shows great potential for sharing these models between ships in a fleet, providing valuable guidance to vessel owners and operators. Using data from multiple ships in a fleet to

train the SLNN may also prove beneficial in the future.

In addition to the successful development and demonstration of this shared-layer structure using three diverse wave data sources, this study also generated unique weight initialization and CV subset partitioning approaches. The weight initialization approach leveraged the PBMPs to begin SLNN training with loss values similar to or better than these underlying PBMPs, which increased training speed and stability for consistent improvement upon them. The CV subset partitioning approach avoided random shuffling (which led to overfitting in preliminary experiments from Chapter V), but still allowed consideration of numerous data subsets for model selection via circular shifts. The SLNN approach developed in this chapter showed great promise for reliable response prediction and improvement upon underlying PBMPs where a traditional NN would be less feasible.

CHAPTER VII

RAO Corrections using Model-Scale Data

While the data-driven approaches of Chapters IV, V, and VI proved effective for vessel motions (and structural responses, as mentioned in Chapter IV), it was difficult to trace the contributions of different uncertainty sources (i.e. wave data, seakeeping model, onboard instruments, etc.) to the underlying error that was being corrected. In theory, these prediction approaches accounted for inadequacies in the numerical models used to generate the RAOs; however, the discrepancies between predicted and measured values also resulted from uncertainties in the wave conditions, which may be significant as shown in the wave data study of Chapter III. In addition to uncertainty in the characteristic wave parameters incorporated into the prior models, other factors, such as directional spread and spectral shape, also affected the accuracy of the framework's output. To better understand the cause of discrepancies between the initial twin framework's predictions and real-world measurements, the focus of this chapter was correction of the RAOs used by the framework. In contrast to the previous data-driven approaches, which corrected the response amplitudes (or variances) derived from the response spectrum, these new correction models seek to make corrections in the spectral density and RAO space as a function of wave frequency, ω_0. An approach for RAO correction using model-scale data is outlined in the following sections.

7.1 Model-Scale Data RAO Correction Formulation

As shown in Figure 7.1, in addition to ω_0, the other two input variables to the RAO correction model were the vessel speed, U, and the heading of the vessel relative to the incoming waves, μ. The output of the correction model was an RAO correction factor, $CF(\omega_0, U, \mu)$, which is defined in Equation 7.1 as the square root of the ratio of measured to predicted pseudo-response spectral densities, $S_M^+(\omega_0, U, \mu)$ and $S_P^+(\omega_0, U, \mu)$, respectively.

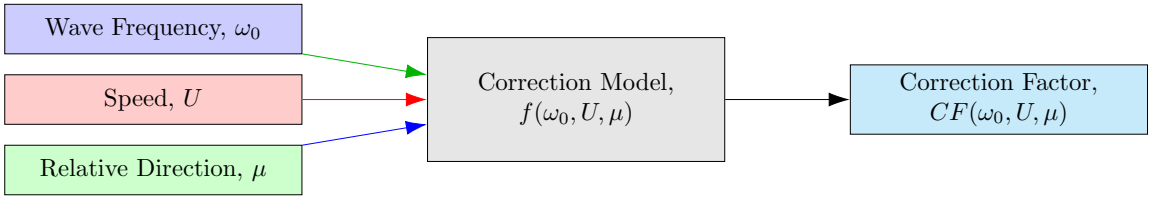

Figure 7.1: Summary of the RAO correction factor, $CF(\omega_0, U, \mu)$, approach formulation.

$$CF(\omega_0, U, \mu) \equiv \sqrt{\frac{S_M^+(\omega_0, U, \mu)}{S_P^+(\omega_0, U, \mu)}} \equiv \frac{|RAO_M(\omega_0, U, \mu)|}{|RAO_P(\omega_0, U, \mu)|} \tag{7.1}$$

As specified in Equation 7.1, under the assumptions of stationarity and linearity, and given unidirectional seas with a perfectly known wave spectrum, this correction factor is equivalent to the ratio of measured to predicted RAO magnitudes, $|RAO_M(\omega_0, U, \mu)|$ and $|RAO_P(\omega_0, U, \mu)|$, respectively. Of course, these assumptions warrant further verification, especially when using full-scale training data from an operational ship. Given a properly trained correction model and the validity of these assumptions, future response predictions could be calculated using the corrected RAOs, $RAO_C(\omega_0, U, \mu)$, with the corrected magnitude defined in Equation 7.2.

$$|RAO_C(\omega_0, U, \mu)| = CF(\omega_0, U, \mu) \times |RAO_P(\omega_0, U, \mu)| \tag{7.2}$$

Note that unlike the approaches of Chapters IV, V, and VI, this approach was not dependent on the significant wave height. This input variable omission was purposeful to maintain the linearity of this RAO correction model; i.e., the correction models were only based on the RAOs' input variables. Further models to address non-Gaussianity of the responses, such as those discussed in *Zhu and Collette* (2017), could be investigated in the future.

The correction factor formulation introduced above relied upon construction and training of a correction "surface." This surface was defined in four dimensions: ω_0, U, μ, and the output correction factor, $CF(\omega_0, U, \mu)$. While the LS, NN, and SLNN approaches were effective, these models were parametric, meaning the number of free parameters was user-defined. As detailed in Chapter V, a physical basis was incorporated into these data-driven models by including the initial twin framework's predictions of response standard deviations, or PBMPs, into the input data matrix. Also mentioned previously, *Weymouth and Yue* (2013) incorporated a physical basis into machine learning models to predict head-seas RAOs from model-scale experiments of the Wigley hull. Their nonparametric formulation of the physics-based model involved complimenting a generic learning model kernel function of input data with a kernel function of the physical basis. As described in *Bishop* (2006), kernel methods can be applied when the vector of inputs, $\mathbf{x}$, only enters an algorithm in the form of scalar products. *Weymouth and Yue* (2013) reduced the physical basis to a 3×1 vector of the physical model's output for a given input and the model's output one input data increment before and after the given value, which was the result of a Taylor series expansion of the physical function. While this physics-based approach was very intriguing for these applications, and similar approaches may be worth investigating in the future, it was too intensive for this investigation, which had limited training data, as described in Section 7.3; however, *Weymouth and Yue* (2013) did demonstrate the potential advantages of non-parametric kernel methods

for this work.

Several kernel methods were considered for the RAO corrections including Gaussian process regression (GPR), support vector machine regression (SVMR), and relevance vector machine regression (RVMR). One advantage of GPR and RVMR over SVMR is that they provide posterior probabilistic outputs, which is desirable for uncertainty quantification in a reflective twin framework (*Bishop*, 2006). GPR is memory-based, meaning it requires storage of all training data points, which can be a computational burden; however, it is possible that training data reduction techniques such as input dimension filtering based on automatic relevance determination (ARD), which is described in Section 7.2, could be used to improve tractability and computation speed (*Rasmussen and Williams*, 2006). While SVMR has not been selected as a candidate for this work due to its lack of probabilistic output, both SVMR and RVMR address the issue of storage by only keeping some of the training data. While RVMR and other probabilistic approaches may be valid candidates for future work, especially given denser data sets, GPR was the focus here. An evaluation of the formulation given in Figure 7.1 was completed using model-scale seakeeping data relevant to the *R/V Knorr* of Chapter IV to train GPR correction models, with results presented in the following sections.

7.2 Gaussian Process Regression (GPR)

Gaussian Process Regression (GPR) was selected as the model for RAO correction. The GPR methodology presented in this section follows that of *Rasmussen and Williams* (2006), and the GPR model construction and training conducted in this work utilized the `Python` package `GPy` (*GPy*, since 2012). As specified in *Rasmussen and Williams* (2006), a Gaussian process is a collection of random variables, any finite number of which have a joint Gaussian distribution. As such, the Gaussian process is defined by a mean function, $m(\mathbf{x})$, and a covariance function, $k(\mathbf{x}, \mathbf{x}')$. The mean

function is often set to zero for GPR, which was the case in this work, and the form of the covariance, or kernel, function is user-specified with parameters that are learned from the train data. Each real-world observation of vessel motion, y_i, available to train the GPR model contains noise; therefore, as outlined in Equation 7.3, y_i can be idealized as the "true" function to be modeled, $f(\mathbf{x}_i)$, plus a Gaussian noise term, ϵ (*Rasmussen and Williams*, 2006).

$$y_i = f(\mathbf{x}_i) + \epsilon \tag{7.3}$$

The Gaussian noise terms for each of the n observations contained in $\mathbf{y}$ are assumed to be independent and identically distributed with variance σ_n^2. The kernel selected for exploration with this model was a radial basis function (RBF) (a.k.a. squared exponential) kernel, which is defined for two input vectors, $\mathbf{x}_i$ and $\mathbf{x}_j$ in Equation 7.4 (*Rasmussen and Williams*, 2006).

$$\text{cov}(f(\mathbf{x}_i), f(\mathbf{x}_j)) = k(\mathbf{x}_i, \mathbf{x}_j) = \sigma_f^2 \exp\left(-\frac{1}{2}(\mathbf{x}_i - \mathbf{x}_j)^T \text{diag}(\boldsymbol{\ell})^{-2}(\mathbf{x}_i - \mathbf{x}_j)\right) + \sigma_n^2 \delta_{ij} \tag{7.4}$$

In Equation 7.4, σ_f^2 is the signal variance, a tunable hyperparameter used to fit the train data, and $\boldsymbol{\ell}$ is a vector of d lengthscales, which are also tunable hyperparameters, corresponding to the d input dimensions contained in $\mathbf{x}_i$ or $\mathbf{x}_j$. Using a separate lengthscale for each dimension rather than a single value for all dimensions is referred to as automatic relevance determination (ARD), which can be used to reduce model complexity by revealing minimally relevant input dimensions that can be subsequently removed from the model (*Rasmussen and Williams*, 2006). As mentioned previously, the observed vessel motions stored in $\mathbf{y}$ contain noise; therefore, the observation noise variance, σ_n^2, was added to the covariance function as shown in Equation 7.5, where δ_{ij} is the Kronecker delta function. Equation 7.6 gives the matrix equivalent of this covariance function, where matrix entry $K_{ij}(\mathbf{X}, \mathbf{X})$ is equal to $k(\mathbf{x}_i, \mathbf{x}_j)$, and $\mathbf{I}$ is the

identity matrix.

$$\text{cov}(y_i, y_j) = k(\mathbf{x}_i, \mathbf{x}_j) + \sigma_n^2 \delta_{ij} \tag{7.5}$$

$$\text{cov}(\mathbf{y}) = K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} \tag{7.6}$$

As a result of the above formulation, the hyperparameters tuned in training the GPR model were one lengthscale for each input dimension, each contained in $\boldsymbol{\ell}$, the RBF signal variance, σ_f^2, and the Gaussian noise variance, σ_n^2. To tune the hyperparameters of a GPR model, the goal is to maximize the marginal likelihood, $p(\mathbf{y}|\mathbf{X})$, which is marginalized over the function values $\mathbf{f}$, as specified in Equation 7.7 (*Rasmussen and Williams*, 2006).

$$p(\mathbf{y}|\mathbf{X}) = \int p(\mathbf{y}|\mathbf{f}, \mathbf{X}) p(\mathbf{f}|\mathbf{X}) d\mathbf{f} \tag{7.7}$$

In Equation 7.7, $p(\mathbf{y}|\mathbf{f}, \mathbf{X})$ is the likelihood, which is a factorized Gaussian distribution with mean $f(\mathbf{X})$ and covariance matrix $\sigma_n^2 \mathbf{I}$, and $p(\mathbf{f}|\mathbf{X})$ is the prior, which is a Gaussian distribution with mean $\mathbf{0}$ and covariance matrix $K(\mathbf{X}, \mathbf{X})$. This yields the log marginal likelihood, defined in Equation 7.8, which is maximized to determine the values of the hyperparameters for the trained model.

$$\log p(\mathbf{y}|\mathbf{X}) = -\frac{1}{2}\mathbf{y}^T (K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I})^{-1} \mathbf{y} - \frac{1}{2} \log |K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I}| - \frac{n}{2} \log 2\pi \tag{7.8}$$

The Limited-memory BFGS optimizer of the `GPy` package was used to determine these hyperparameters through maximization of Equation 7.8.

After training the model using the observed values of $\mathbf{y}$ and corresponding input variables stored in $\mathbf{X}$, the joint distribution of $\mathbf{y}$ and the predicted values, $\mathbf{f}_*$, given test input matrix, $\mathbf{X}_*$, is defined in Equation 7.9 (*Rasmussen and Williams*, 2006).

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{f}_* \end{bmatrix} \sim \mathcal{N}\left(\mathbf{0}, \begin{bmatrix} K(\mathbf{X}, \mathbf{X}) + \sigma_n^2 \mathbf{I} & K(\mathbf{X}, \mathbf{X}_*) \\ K(\mathbf{X}_*, \mathbf{X}) & K(\mathbf{X}_*, \mathbf{X}_*) \end{bmatrix} \right) \tag{7.9}$$

Based on the formulation presented in Equation 7.9, the conditional distribution for making GP predictions, $\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{X}_*$, can be derived, which is defined in Equations 7.10-7.12.

$$\mathbf{f}_*|\mathbf{X}, \mathbf{y}, \mathbf{X}_* \sim \mathcal{N}(\overline{\mathbf{f}_*}, \text{cov}(\mathbf{f}_*)) \tag{7.10}$$

$$\overline{\mathbf{f}_*} = K(\mathbf{X}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2\mathbf{I}]^{-1}\mathbf{y} \tag{7.11}$$

$$\text{cov}(\mathbf{f}_*) = K(\mathbf{X}_*, \mathbf{X}_*) - K(\mathbf{X}_*, \mathbf{X})[K(\mathbf{X}, \mathbf{X}) + \sigma_n^2\mathbf{I}]^{-1}K(\mathbf{X}, \mathbf{X}_*) \tag{7.12}$$

In Equations 7.10-7.12, $\overline{\mathbf{f}_*}$ is the predictive mean, which is a vector of the GPR model's best estimates given new input data points contained in $\mathbf{X}_*$, and $\text{cov}(\mathbf{f}_*)$ is the predictive variance, which is a vector of the variance corresponding to the input data points. While the discussion presented in the following section focuses solely on the RAO correction model's predictive means due to the preliminary nature of this study, predictive variance is considered in Chapter VIII for RAO correction using full-scale data.

7.3 GPR Corrections Using Model-Scale Data

Data were acquired from model-scale seakeeping experiments performed at the Naval Surface Warfare Center Carderock Division (NSWCCD) for the *R/V Melville*, which is a sister ship of the *R/V Knorr*. These model-scale experiments are described in *Minnick et al.* (2012) and included tests in both regular and irregular unidirectional waves, with the irregular tests performed in up to three different sea states. The provided data included RAOs that were calculated using measured response spectra and wave spectra, which approximately followed a Bretschneider spectrum as defined in Equations 2.1-2.3. The goal of the exploration presented here was to use the model-scale experimental data to train a GPR model for prediction of RAO correction factors, $CF(\omega_0, U, \mu)$. This correction approach was then tested by recalculating the

R/V Knorr's uncorrected, full-scale responses from the 2013 cruise of *Schirmann et al.* (2019a) and Chapter IV using the updated RAOs within the initial twin framework. Unlike the LS and NN approaches of *Schirmann et al.* (2019a) and Chapter IV, this GPR approach could be tested using both unidirectional and multidirectional NWW3 data. Also in contrast to the previous data-driven models, this approach did not combine degrees of freedom (e.g. heave and pitch) as input to the same correction model. For the results presented here, this investigation focused solely on vessel pitch.

Model-scale RAOs were available for select relative headings, μ, in both regular and irregular unidirectional waves, at two (full-scale) speeds: 8 and 12 knots (*Minnick et al.*, 2012). The availability of data at only two speeds limited this investigation to a relatively small region of the *Knorr*'s 2013 cruise operating window, with only full-scale data at speeds between 7.5 knots and 12.5 knots relevant for testing. The lack of RAO data between these two speeds also prevented development of the GPR correction surface in all four of the previously mentioned dimensions (ω_0, U, μ, and $CF(\omega_0, U, \mu)$). Instead, one correction factor surface was trained in three dimensions (ω_0, μ, and $CF(\omega_0, \mu)$) at each speed. Linear interpolation in U was then used to correct the RAOs between 8 and 12 knots, which were defined for all relative headings spaced in 5° increments at speeds of 9, 10, and 11 knots.

The model-scale data were also relatively sparse in the μ dimension, with data only provided in head (180°), bow-quartering (135°), beam (90°), stern-quartering (45°), and following (0°) seas. Furthermore, some of the provided RAO data at these headings excluded regular wave test results. Preliminary experimentation with GPR models for the correction factor using these data revealed issues with determination of an appropriate lengthscale for the μ dimension, likely due to data sparsity. In order to determine a lengthscale for this dimension that was informed by physics, separate GPR surfaces were trained to model the uncorrected RAOs from `SHIPMO.BM` using the same two input variables, ω_0 and μ, at both 8 and 12 knots. Although the

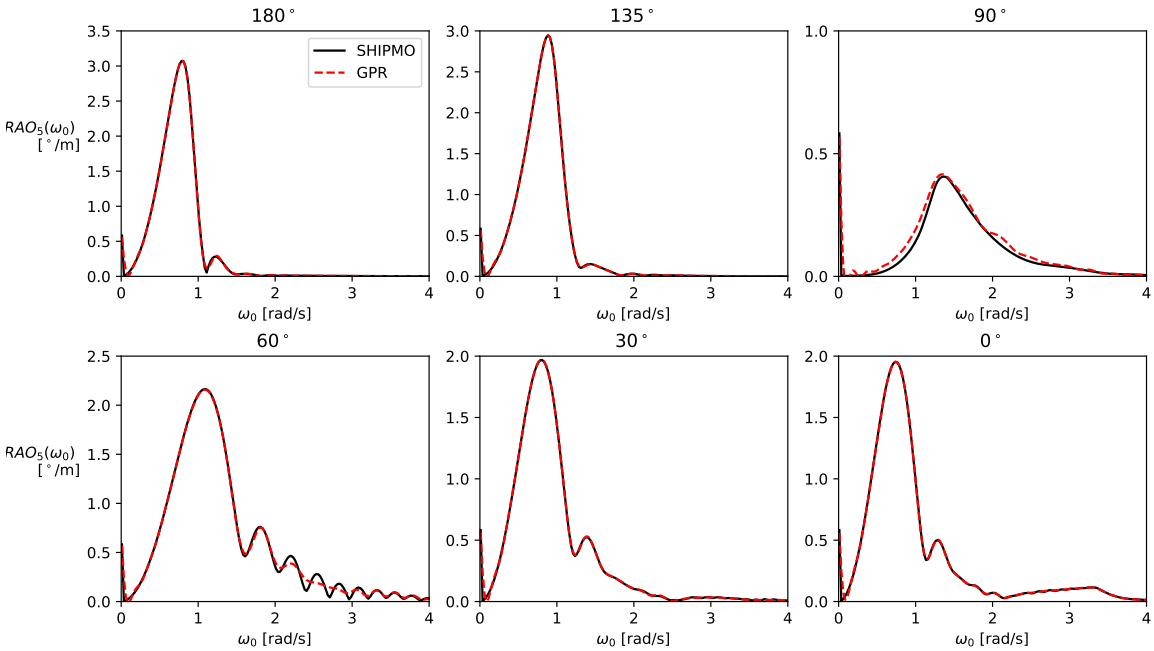

Figure 7.2: Examples of the GPR model's fit to `SHIPMO.BM` RAOs using a μ dimension lengthscale equivalent to 10° before normalization.

optimizer used in training the GPR often yielded different lengthscales each time it was run, likely due to numerical instability in the training process, it was determined that a μ lengthscale of 10° (before normalization) yielded a relatively good fit to all of the `SHIPMO.BM` RAOs at each speed. An example of this fit for select directions at 8 knots is shown in Figure 7.2. As shown in Figure 7.2, the fit is very strong for most of the RAOs, with subtle deviations for directions of 90° and 60°. Similar performance was seen for all of the RAOs, each separated by a 5° heading increment, at both speeds. This 10° lengthscale, which was shown to have a physical basis through experimentation, was then constrained fixed in constructing and training the GPR model for correction factors using model-scale data. After constraining the μ dimension lengthscale, three hyperparameters remained for optimization: the ω_0 dimension lengthscale, the noise term in the RBF (σ_f^2) and the Gaussian noise term (σ_n^2).

One further complexity of the experimental RAO data were that the irregular

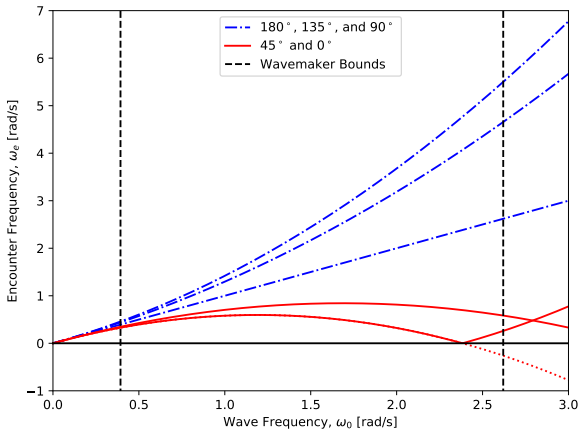

Figure 7.3: Plot of encounter frequency, ω_e, versus wave frequency, ω_0, showing the wavemaker bounds used to filter the irregular wave model-scale data.

wave results were calculated as a function of encounter frequency, ω_e. While a single wave frequency could be recovered in head, bow-quartering, and beam seas (180°, 135°, and 90°), in following and stern-quartering seas (45° and 0°), up to three wave frequencies could yield the same encounter frequency, as discussed in *Lloyd* (1989). This complexity is demonstrated mathematically in Equation 7.13 and graphically in Figure 7.3.

$$\omega_0 = \frac{g}{2U\cos(\mu)}\left(1 \pm \sqrt{1 - \frac{4(\pm\omega_e)}{g}U\cos(\mu)}\right) \tag{7.13}$$

Note that $\pm\omega_e$ is used in Equation 7.13 because negative ω_e values could not be distinguished from positive frequencies in the experimental data, which is reflected in Figure 7.3 where the 0° curve is plotted as the absolute value of the encountered frequency, with a dotted line showing its signed trajectory. The regular wave test data were not affected by this issue because the singular wave frequency of each test was known. To avoid discarding all of this potentially valuable data from the irregular wave tests in stern-quartering and following seas, a logic-based filtering approach was applied. As shown in Figure 7.3, upper and lower wave frequency bounds were used to filter the candidate wave frequencies for a given encounter frequency based on the wavemaker specifications provided in *Minnick et al.* (2012). Next, a logical test was

applied in which data points were kept only if one of the remaining candidate wave frequencies corresponded to a `SHIPMO.BM` RAO value that was at least 10 times greater than all of the other candidates' RAO values. In the case that a data point was kept to train the GPR model, it was assigned the wave frequency corresponding to this dominating RAO value. While this logic-based approach is imperfect and relies upon the assumption that the relative RAO magnitude between different wave frequencies predicted by `SHIPMO.BM` holds in practice, it was deemed necessary to avoid throwing out nearly all of the following seas experimental data. Access to more regular wave test data in stern-quartering and following seas would be the best way to address this issue. A more rigorous filtering approach, made possible by a larger pool of data, is discussed for GPR RAO correction using full-scale data in Chapter VIII.

To make the correction factor surface more closely model a zero-mean process as specified in Section 7.2, the transformation presented in Equation 7.14 was applied to the i^{th} correction factor from the model-scale train data, CF_i, to yield the GPR train data targets, y_i.

$$y_i = \frac{1}{CF_i} - 1 \tag{7.14}$$

Also note that the input variables, ω_{0i} and μ_i, were also normalized by 4.885 rad/s, the maximum frequency of the `SHIPMO.BM` RAOs, and 180°, respectively.

After training the GPR correction model, the `SHIPMO.BM` RAOs for all headings and speeds of 8, 9, 10, 11, and 12 knots were updated using the correction factors, and the initial twin framework was used to calculate updated response predictions for the *R/V Knorr*'s 2013 cruise, once again using unidirectional NWW3 wave data and a spreading function as presented in Chapter IV. These response predictions were also calculated using multidirectional wave data from the NWW3 Production Hindcast, which were available in one-hour time increments rather than the three-hour increments of the unidirectional data. The multidirectional wave data also included defined directional spreads for each wave partition, which were implemented using

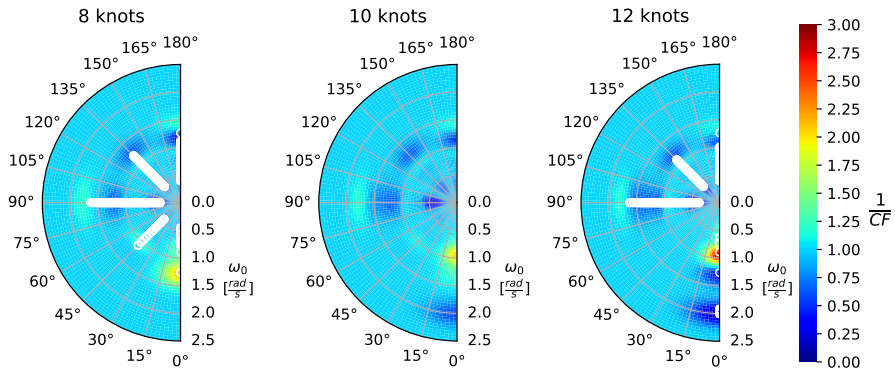

Figure 7.4: Polar plots of the inverse RAO correction factors, $1/CF$, from the trained GPR model at 8, 10, and 12 knots. The white circles represent input data coordinates where model-scale data were available.

spreading functions.

7.3.1 Model-Scale Data GPR Results

The inverse correction factor, $1/CF$, surfaces yielded by the trained GPR model using the model-scale, experimental data for 8 and 12 knots are shown in Figure 7.4, which also shows the linearly interpolated surface at 10 knots. The inverse of the correction factor was plotted to prevent large CF values from dominating the plots. The 8 and 12 knot polar plots of Figure 7.4 also include white circles at the ω_0 and μ coordinates where a training data point was used. While the results are plotted in polar coordinates, these GPR models were constructed in a three-dimensional Cartesian space, meaning the CF values at lower frequencies did not depend more heavily on the CF values of other relative directions than those at higher frequencies. As shown in Figure 7.4, the GPR model for the correction factors returned to a value of 1.0 in regions of the input space where training data were unavailable. This means that RAOs that were not well represented in the experimental data remained unchanged. The spread in the μ dimension resulted from the fixed 10° lengthscale that was learned from the `SHIPMO.BM` GPR model. The updated RAOs from this correction approach are compared to the training data RAOs from the model-scale

experiments and the original SHIPMO.BM RAOs in Figures 7.5 and 7.6 for 8 and 12 knots, respectively.

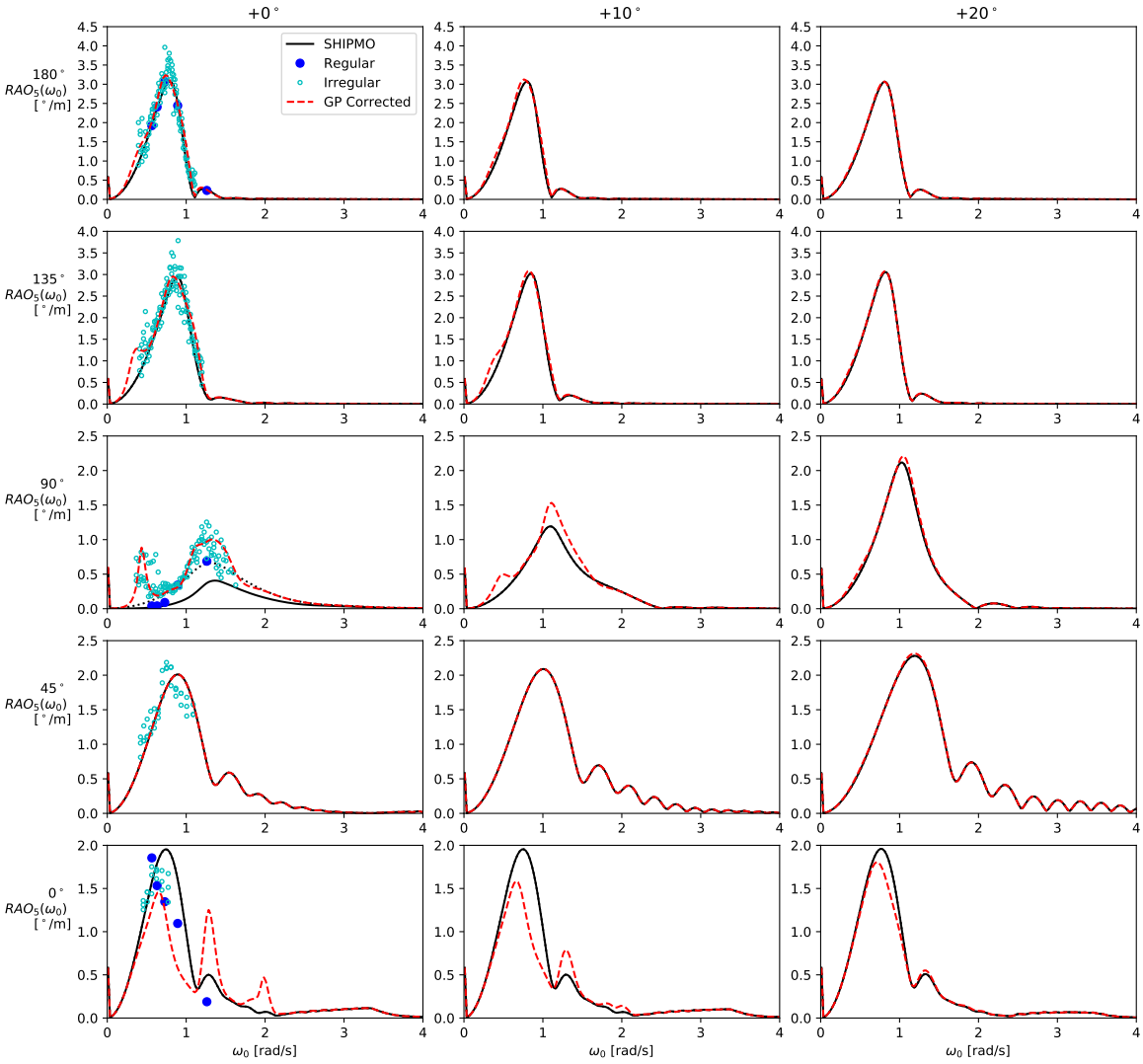

Figure 7.5: Examples of the GPR model's fit to the model-scale, experimental RAOs at 8 knots for all five headings, as well as the RAOs 10° and 20° greater than each heading.

As shown in Figures 7.5 and 7.6, the GPR updated RAOs deviated from the SHIPMO.BM RAOs at wave frequencies where experimental data were available, but returned to the original SHIPMO.BM RAOs in regions where data were not available. The updated RAOs at headings 10° and 20° greater than each of the experimental directions are also included in Figures 7.5 and 7.6. These additional plots were included

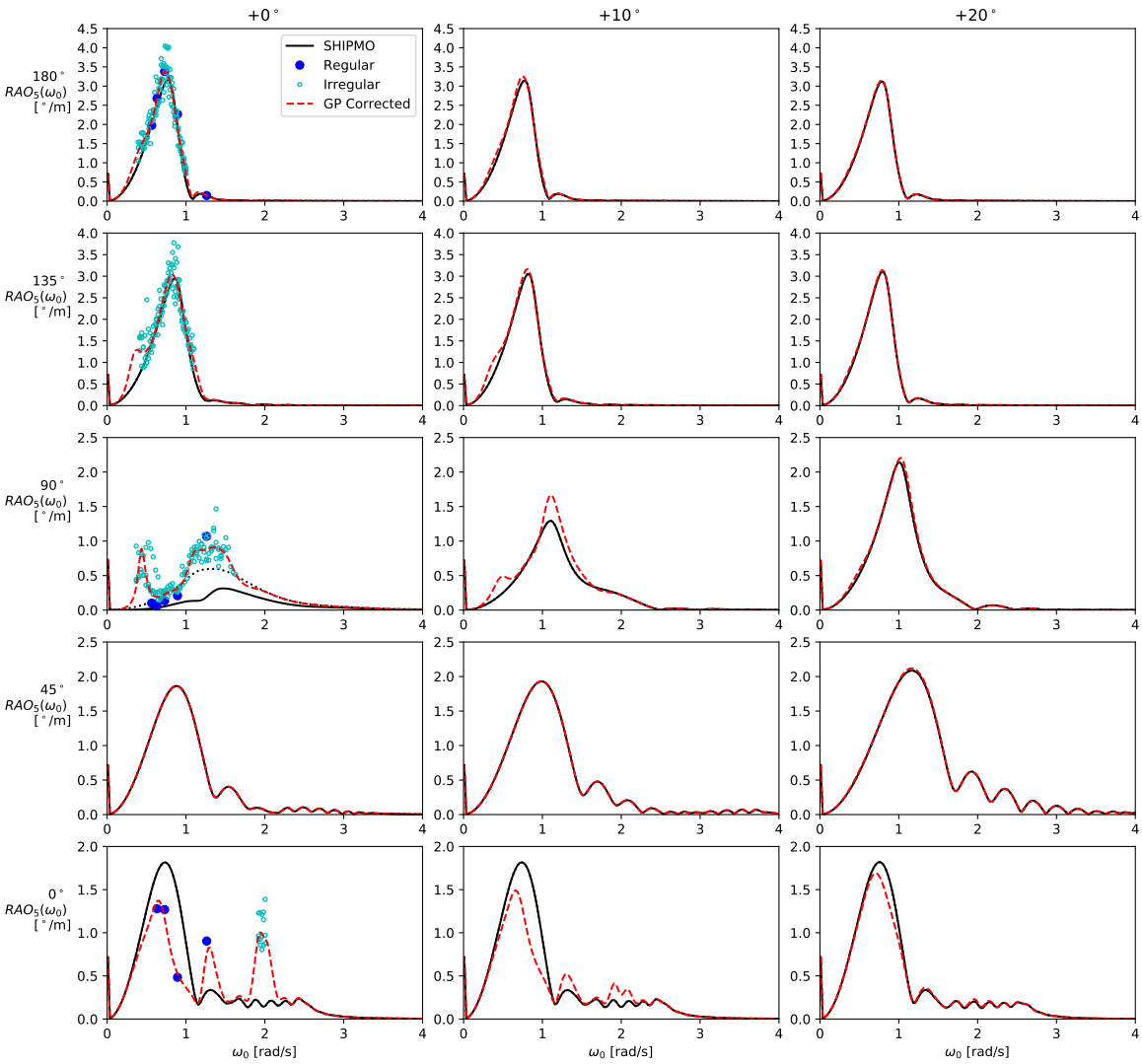

Figure 7.6: Examples of the GPR model's fit to the model-scale, experimental RAOs at 12 knots for all five headings, as well as the RAOs 10° and 20° greater than each heading.

to show that the GPR-updated RAOs also returned to the original `SHIPMO.BM` values as the distance in μ from experimental data increased at a pace defined by the fixed μ lengthscale. Note that this approach did not distinguish between regular and irregular wave data points in training the model, which could be incorporated in future work. It should also be noted that the experimental data were sparse and imperfect, and some of the experimental RAOs may not have resulted from the hydrodynamic properties of the hull form that would propagate to full-scale. As such, it is important

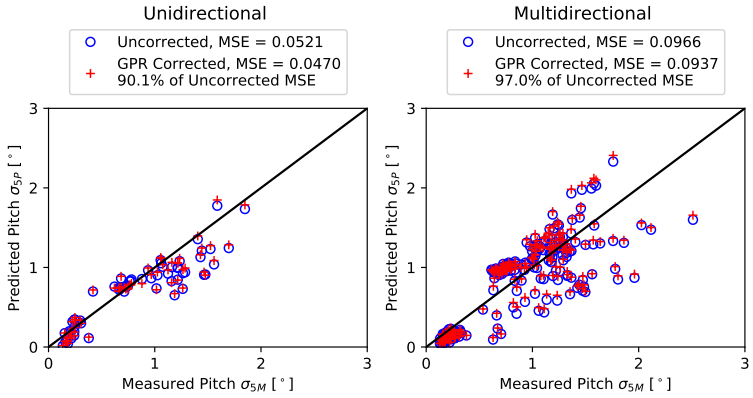

Figure 7.7: Predicted versus measured pitch predictions using GPR-corrected RAOs with unidirectional and multidirectional wave data for speeds between 7.5 and 12.5 knots. The black lines with slope 1.0 represent perfect agreement between predicted and measured values.

to consider that these results could be used as a single step in a correction process that incorporates full-scale data for further correction.

The initial twin framework's updated pitch response predictions using the GPR-corrected RAOs with both unidirectional and multidirectional wave data at speeds between 7.5 and 12.5 knots are presented in Figure 7.7. There are approximately three times more data points for the multidirectional case because data were available in one-hour time increments. The speed and heading standard deviation filters discussed in Section 4.1.2 were also used to filter these results. As shown in the legend of Figure 7.7, the MSE of the full-scale pitch predictions was reduced slightly for both the unidirectional and multidirectional cases. Visually, it is also evident that the effects of the updated RAOs were very minor. An important consideration regarding these results is that while the experimental data used to train the GPR model for RAO correction had a known wave spectrum shape, the full-scale predictions of the initial twin framework were still calculated under the assumption of a Bretschneider spectrum for each wave partition. While the subtlety of this correction approach's effects on the response predictions may be due to the limited model-scale training data available, it may also suggest that uncertainty in the shape of the encountered wave

spectrum had a very significant effect on the accuracy of these response predictions.

7.3.2 Model-Scale Data GPR Conclusions

While further model-scale testing of the *R/V Melville* with a wider array of speeds and headings was planned at the University of Michigan's Marine Hydrodynamics Laboratory, logistical challenges, amplified by the global pandemic, prevented the plan's execution. Nevertheless, model-scale experimental data should undoubtedly be considered in real-world reflective twin implementation. Given the large pool of available full-scale data outlined in Chapter V, further study of RAO corrections in this body of work used full-scale data from the *R/V Neil Armstrong*, as outlined in the following chapter. The insights and experience gained from these experiments with GPR correction allowed for development of more rigorous data filtering and GPR model construction approaches in the following chapter.

CHAPTER VIII

RAO Corrections using Full-Scale Data and Two-Stage Model Demonstration

This chapter details GPR RAO correction using full-scale data from the *R/V Neil Armstrong*. The ability to train RAO correction models using full-scale data would mitigate or eliminate the need for expensive model-scale tests to measure RAOs, as described in the following discussion. The GPR-corrected RAOs produced in this study were then used in a model demonstration with two distinct stages:

1. GPR-corrected RAOs were used in the initial twin framework to generate heave, pitch, and roll PBMPs for each of the three wave data sources of Chapter VI.

2. These PBMPs, generated with GPR-corrected RAOs, were used to construct, train, and test SLNNs for the three wave data sources of Chapter VI.

For both stages of the demonstration, model performance was directly compared to the Chapter VI results to evaluate the potential benefits of incorporating GPR RAO correction alone (stage one) or as a pre-processing step for the SLNN (stage two).

As discussed in Chapter VII, knowledge regarding the wave spectrum shape is necessary for accurate RAO calculation from the measured response spectrum. To address this need, the present study used CDS ERA5 2D wave spectra, which were also employed as a wave data source in Chapter VI (*Copernicus Climate Change*

Service [C3S], 2021; *Hersbach et al.*, 2020). A study conducted by *Nielsen et al.* (2021) also employed ERA5 2D wave spectra for pitch RAO tuning. In contrast to this study, the CDS wave data were used in an optimization-based approach, and the underlying pitch RAOs were calculated using semi-analytical transfer functions rather than 2D strip theory. Furthermore, *Nielsen et al.* (2021) only considered data from a single voyage.

As mentioned in Chapter VII, even given known spectral shapes, challenges arise for RAO calculation in stern-quartering and following sea conditions when multiple wave frequencies correspond to the same encounter frequency experienced by the moving ship. Rather than discard all of these valuable wave data, a robust filtering approach was developed to identify specific sea conditions and encounter frequencies when full-scale measurements could be assigned to a single wave frequency at a given speed and heading with relatively high confidence for calculation of a corresponding RAO. Unlike the limited operating conditions available from the model-scale experiments of Chapter VII, this approach allowed for GPR model training across an array of speeds and headings from the AR Train dataset. Additionally, the developed filtering approach did not require the use of 2D strip theory RAOs for assignment of response energy to specific wave frequencies in stern-quartering and following seas, as was done in Chapter VII. The results of this GPR RAO correction approach and the subsequent SLNN demonstration quantify the effectiveness of these data-driven approaches separately and jointly for improved response prediction using the full, real-world datasets collected aboard the *Neil Armstrong* and *Sally Ride*.

8.1 Methodology

This section details the developed wave data filtering approach and the use of these wave data for measured RAO calculation and subsequent GPR RAO correction. This section also details the use of these GPR-corrected RAOs in the initial twin framework

for PBMP generation, and employment of these GPR PBMPs in the SLNN.

8.1.1 Wave Data Selection

As discussed in Chapter VII, for a vessel with non-zero speed in irregular, following to beam seas (i.e., wave direction relative to ship heading $\mu \in [0°, 90°)$), assigning a spectral density value from the vessel's encountered response spectrum to a single wave frequency is a difficult (sometimes impossible) task. Furthermore, ensuring that wave conditions are approximately unidirectional is another challenge that arises for RAO correction when using full-scale data; real-world sea conditions are frequently short-crested to varying degrees. This section details a robust filtering process to select spectral wave data for RAO calculation when there is high confidence that seas were relatively unidirectional and a given spectral density from the vessel's response spectrum can be attributed to a single wave frequency. Note that this approach considers unidirectional conditions at the wave-frequency level, which prevents unnecessary discardment of valuable spectral wave data at one wave frequency based solely on significant directional spreading at another wave frequency. For example, if a higher wave frequency has an unacceptable level of directional spreading for reliable RAO calculation, it does not necessarily mean that a lower wave frequency also has unacceptable spreading.

As described in Chapter VI, the 2D wave spectra for a given time and location were discretely defined for 24 evenly-spaced directions on [0°, 360°) at 30 frequencies (*Copernicus Climate Change Service [C3S]*, 2021; *Hersbach et al.*, 2020). The spacing of these frequencies is defined in Equation 8.1.

$$f_{CDS}(n) = 0.0345(1.1)^{(n-1)} \text{ Hz for } n = 1, 2, ..., 30 \tag{8.1}$$

Converting to radians, these frequencies therefore ranged from a minimum value,

$\min(\omega_{CDS})$, of approximately 0.2168 rad/s to a maximum value, $\max(\omega_{CDS})$, of approximately 3.4386 rad/s.

Similar to the approach described in Chapter VI, the 2D wave spectra were linearly interpolated between the 8 corners in latitude, longitude, and time space (as defined in Chapter VI). For discrete CDS wave direction j and frequency k, the interpolated 2D wave spectral density value, $S_{CDS}(\theta_j, \omega_k)$, was calculated using the scale factor SF_i (defined in Equation 6.2), and the 2D wave spectral density value, $S_{CDS,i}(\theta_j, \omega_k)$, for each corner i, as shown in Equation 8.2.

$$S_{CDS}(\theta_j, \omega_k) = \sum_{i=1}^{8} SF_i \times S_{CDS,i}(\theta_j, \omega_k) \tag{8.2}$$

The discrete, interpolated values of $S_{CDS}(\theta_j, \omega_k)$ were then used to define the 2D wave spectrum considered for RAO calculation, as described in the following paragraph.

Let the discrete wave spectral densities of $S_{CDS}(\theta_j, \omega_k)$, as well as linearly interpolated values (in direction and frequency space) between the discrete wave directions and frequencies of $S_{CDS}(\theta_j, \omega_k)$, define continuous 2D CDS spectral density function $S_{CDS}(\theta, \omega)$. Given that this CDS spectral density function was only defined between the specified frequency bounds, the full 2D spectral density function used moving forward, $S(\theta, \omega)$, is defined in Equation 8.3.

$$S(\theta, \omega) = \begin{cases} 0, & \omega \leq 0 \\ S_{CDS}(\theta, \min(\omega_{CDS})), & 0 < \omega \leq \min(\omega_{CDS}) \\ S_{CDS}(\theta, \omega) & \min(\omega_{CDS}) < \omega \leq \max(\omega_{CDS}) \\ 0, & \omega > \max(\omega_{CDS}) \end{cases} \tag{8.3}$$

Note that the spectral density for $0 < \omega \leq \min(\omega_{CDS})$ was set to $S_{CDS}(\theta, \min(\omega_{CDS}))$ rather than 0 to avoid discarding any existent low-frequency energy content.

The continuous 2D spectral density function defined in Equation 8.3 was used for six wave data filters described in the following sections. These filters addressed whether or not to retain a given wave frequency, ω, and its associated spectral density for RAO calculation and GPR model training. Note that the "0" subscript appended to wave frequency variable ω in previous chapters was omitted in this section to simplify notation. Also note that due to the linear nature of response prediction with RAOs, as a pre-processing step, it was deemed necessary to remove any AR Train windows with interpolated significant wave heights greater than two meters from consideration. This two-meter threshold was selected to conservatively ensure that sea states were not severe, which may prompt nonlinear responses. While nonlinear responses are certainly of interest for improved response prediction, these instances were deferred for later correction with the SLNN. Note that the specific wave frequencies provided as candidates to these six filters are included in later discussion.

8.1.1.1 Filter 1: 1D Spectral Density Threshold

Given the definition of $S(\theta, \omega)$ in Equation 8.3, the 1D spectral density for a given wave frequency, $S(\omega)$, was calculated using Equation 8.4 and trapezoidal integration.

$$S(\omega) = \int_0^{2\pi} S(\theta, \omega) d\theta \tag{8.4}$$

Filter 1 retained wave frequencies where the spectral density criterion outlined in Equation 8.5 was met.

$$S(\omega) \geq 0.01 \ \frac{\mathrm{m^2 s}}{\mathrm{rad}} \tag{8.5}$$

Although subject to user-discretion, the threshold provided in Equation 8.5 was loosely based on the minimum value defined for the 2D wave spectra, $10^{-4} \ \frac{\mathrm{m^2 s}}{\mathrm{rad}}$ (with frequency in Hz, direction in radians). Converting frequency to radians, this minimum 2D spectral density value was equivalent to $\frac{1}{2\pi} \times 10^{-4} \ \frac{\mathrm{m^2 s}}{\mathrm{rad^2}}$. Integrating this

constant value across directions using Equation 8.4 again yielded 10^{-4} $\frac{\mathrm{m^2s}}{\mathrm{rad}}$, the minimum defined value of $S(\omega)$. Given this minimum defined value, the threshold of 0.01 $\frac{\mathrm{m^2s}}{\mathrm{rad}}$ shown in Equation 8.5 was selected, which is 100 times greater than the minimum. Given that wave energy is considered in the denominator for RAO calculation, the purpose of this first filter was to ensure that there was sufficient wave energy present to adequately calculate RAOs. If near-zero wave energy was present at frequency ω, the RAO value could tend to infinity given even a small response spectral density. This behavior would also indicate that the response is not attributable to ω, making it a poor candidate wave frequency for RAO calculation.

8.1.1.2 Filter 2: Directional Spreading Threshold

The mean wave direction at candidate wave frequency ω, $\overline{\theta(\omega)}$, was calculated using Equations 8.6-8.8.

$$A(\omega) = \int_0^{2\pi} \sin(\theta) S(\theta, \omega) d\theta \tag{8.6}$$

$$B(\omega) = \int_0^{2\pi} \cos(\theta) S(\theta, \omega) d\theta \tag{8.7}$$

$$\overline{\theta(\omega)} = \arctan\left(\frac{A(\omega)}{B(\omega)}\right) \tag{8.8}$$

Given mean wave direction $\overline{\theta(\omega)}$ and another arbitrary wave direction, θ, the directional difference from the mean, $\Delta\theta(\theta, \omega)$, was defined as given in Equation 8.9.

$$\Delta\theta(\theta, \omega) = \theta - \overline{\theta(\omega)} \in [-\pi, \pi) \tag{8.9}$$

Rearranging Equation 8.9 yields Equation 8.10 for arbitrary angle θ, another formulation that is more commonly used in the following discussion for consistency with

directional spreading functions.

$$\theta = \overline{\theta(\omega)} + \Delta\theta \tag{8.10}$$

Using the definition presented in Equation 8.10, the mean absolute directional difference at a given wave frequency, $\overline{|\Delta\theta(\omega)|}$, was calculated following Equation 8.11.

$$\overline{|\Delta\theta(\omega)|} = \frac{\int_{-\pi}^{\pi} |\Delta\theta| S\Big(\overline{\theta(\omega)} + \Delta\theta, \omega\Big) d(\Delta\theta)}{S(\omega)} \tag{8.11}$$

The actual implementation of Equation 8.11 in this study, which used summation of discrete $\Delta\theta$ values, is explicitly defined in Equation 8.12.

$$\overline{|\Delta\theta(\omega)|} = \frac{\sum_{\Delta\theta} |\Delta\theta| S\Big(\overline{\theta(\omega)} + \Delta\theta, \omega\Big)}{\sum_{\Delta\theta} S\Big(\overline{\theta(\omega)} + \Delta\theta, \omega\Big)}, \quad \Delta\theta = -\pi, -\frac{179}{180}\pi, \ \ldots\ , \frac{179}{180}\pi \tag{8.12}$$

To determine if the wave spectrum was acceptably unidirectional at candidate wave frequency ω, it was necessary to create a comparison with a commonly used wave spreading function. This study employed the cosine-squared spreading function defined in Equation 8.13.

$$D_{cos^2}(\Delta\theta) = \begin{cases} \frac{2}{\pi}\cos^2(\Delta\theta), & \frac{-\pi}{2} \leq \Delta\theta \leq \frac{\pi}{2} \\ 0, & \text{otherwise} \end{cases} \tag{8.13}$$

As required for a valid spreading function, meaning the function preserves energy, note that Equation 8.14 holds true.

$$\int_{-\pi}^{\pi} D_{cos^2}(\Delta\theta)\ d(\Delta\theta) = 1 \tag{8.14}$$

Analogous to Equation 8.11, the mean absolute directional difference for the cosine-squared spreading function, $\overline{|\Delta\theta_{cos^2}|}$ could then be calculated using Equation 8.15, which yields a constant.

$$\overline{|\Delta\theta_{cos^2}|} = \int_{-\pi}^{\pi} |\Delta\theta| D_{cos^2}(\Delta\theta)\ d(\Delta\theta) = \frac{\pi}{4} - \frac{1}{\pi} \tag{8.15}$$

Although Equation 8.15 yields a constant value, discrete computation of $\overline{|\Delta\theta_{cos^2}|}$ was also performed in a similar manner to Equation 8.12 to validate these calculation procedures, as shown in Equation 8.16.

$$\overline{|\Delta\theta_{cos^2}|} = \frac{\sum_{\Delta\theta} |\Delta\theta| D_{cos^2}(\Delta\theta)}{\sum_{\Delta\theta} D_{cos^2}(\Delta\theta)} \approx \frac{\pi}{4} - \frac{1}{\pi}, \quad \Delta\theta = -\pi, -\frac{179}{180}\pi,\ \ldots\ , \frac{179}{180}\pi \tag{8.16}$$

Through comparison to the cosine-squared spreading function, Filter 2 retained wave frequencies where the criterion outlined in Equation 8.17 was met.

$$\overline{|\Delta\theta(\omega)|} \leq \overline{|\Delta\theta_{cos^2}|} \tag{8.17}$$

The purpose of Filter 2 was to remove candidate frequencies where conditions were not sufficiently unidirectional. Note that the cosine-squared spreading function could be replaced with the cosine-fourth spreading function for more stringent filtering in future studies. Also note that the mean absolute directional difference metric could be replaced with a mean-squared directional difference metric; however, as shown in the results of this study, use of the cosine-squared spreading function and mean absolute directional difference metric was adequate for RAO correction here. Candidate wave frequencies which passed these first two filters, based solely in the wave frequency space, were then subject to filters in the encounter frequency space as described in the following steps.

8.1.1.3 Filter 3: Encounter Frequency Nyquist Threshold

Given the ship's speed, U, and relative heading to the mean wave direction, $\mu\left(\overline{\theta(\omega)}\right)$, the encounter frequency, ω_e, corresponding to candidate wave frequency ω, could be calculated as shown in Equation 8.18, where g is the acceleration due to gravity.

$$\omega_e = \omega - \frac{\omega^2 U}{g}\cos\left(\mu\left(\overline{\theta(\omega)}\right)\right) \tag{8.18}$$

The first filter applied in the encounter frequency space is shown in Equation 8.19, which ensured that any wave frequency selected at this stage would have an attainable response spectrum measurement, meaning ω_e was less than the Nyquist frequency, $\omega_{Nyq.}$.

$$\omega_e \leq \pi \text{ rad} = \frac{1}{2}\omega_s \equiv \omega_{Nyq.} \tag{8.19}$$

The value of π rad in Equation 8.19 was calculated using the 2π rad (or 1 Hz) sampling frequency, ω_s, from the motions data measured aboard the *R/V Neil Armstrong*.

8.1.1.4 Filter 4: Encounter Frequency Energy Threshold

Given encounter frequency ω_e corresponding to candidate wave frequency ω, it was then necessary to shift perspectives to the encountered response spectrum space. Given a response spectral density measured at ω_e, which wave frequencies could be responsible? As mentioned previously, in perfectly unidirectional head to beam seas, just one wave frequency, ω_1, could be responsible, which is defined in Equation 8.20.

$$\omega_1(\theta) = \frac{g}{2U\cos\left(\mu(\theta)\right)}\left(1 - \sqrt{1 - \frac{4(\omega_e)U}{g}\cos\left(\mu(\theta)\right)}\right) \tag{8.20}$$

Note that Equation 8.20 is explicitly defined as a function of arbitrary wave direction θ to highlight that in sea conditions that are not perfectly unidirectional, as is typically the case in real-world operation, each heading angle present in the wave spectrum has

a unique wave frequency associated with ω_e. Therefore, even given beam to head sea conditions (i.e., wave direction relative to ship heading $\mu \in [90°, 180°]$), there is still not complete certainty about a single wave frequency's contribution to the response at ω_e, reinforcing the need for this rigorous filtering approach.

In following to beam seas, the two other candidate wave frequencies that arise given arbitrary wave direction θ are defined in Equations 8.21 and 8.22.

$$\omega_2(\theta) = \frac{g}{2U\cos\big(\mu(\theta)\big)}\left(1 + \sqrt{1 - \frac{4(\omega_e)U}{g}\cos\big(\mu(\theta)\big)}\right) \tag{8.21}$$

$$\omega_3(\theta) = \frac{g}{2U\cos\big(\mu(\theta)\big)}\left(1 + \sqrt{1 - \frac{4(-\omega_e)U}{g}\cos\big(\mu(\theta)\big)}\right) \tag{8.22}$$

These three potential wave frequencies for a given wave direction and encounter frequency at 12 knots are demonstrated graphically in Figure 8.1, which is similar to Figure 7.3 of Chapter VII. The multi-colored curves in Figure 8.1 each correspond to a different relative wave heading as indicated by the select annotations included. The portion of each curve that corresponds to each of the three wave frequency options is indicated, as defined in the plot legend. The vertical background lines indicate the CDS wave frequencies defined in Equation 8.1 for reference. The Nyquist frequency considered in Filter 3 is also shown.

As shown in Figure 8.1, perfect beam seas to head seas ($\mu \in [90°, 180°]$)) yield just one wave frequency for each relative direction and encounter frequency. On the contrary, following seas to beam seas ($[0°, 90°)$) yield up to three wave frequencies for each encounter frequency. The third wave frequency, ω_3, corresponds to a negative encounter frequency (i.e., the ship is overtaking slowly-advancing, higher-frequency waves as described in *Lloyd* (1989)). However, from the ship's instrumentation, this negative encounter frequency cannot be distinguished from a positive encounter frequency; only the absolute value of this negative value can be recovered. As such, the

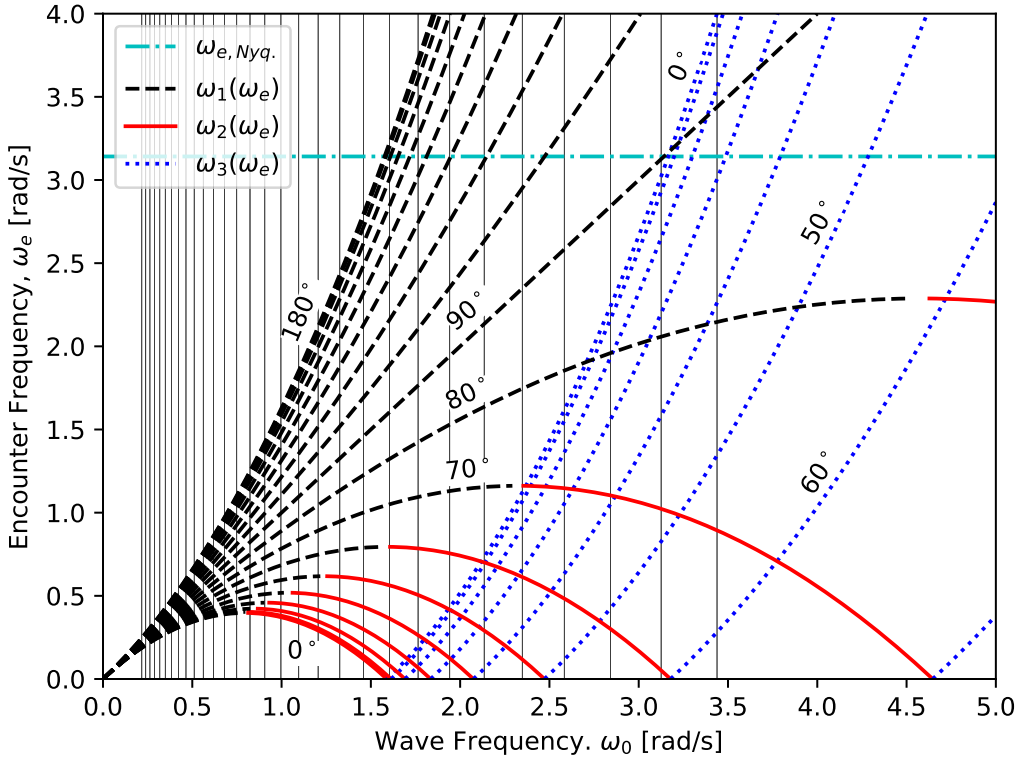

Figure 8.1: Plot of encounter frequency versus wave frequency for select relative headings at 12 knots. The portion of each curve that corresponds to each of the three wave frequency options given in Equations 8.20, 8.21, and 8.22 is indicated. The vertical background lines indicate the CDS wave frequencies for reference. The Nyquist frequency considered in Filter 3 is also shown.

transition from ω_2 to ω_3 demonstrated in Figure 8.1 is a non-smooth "bounce" off the axis where $\omega_e = 0$ rad.

In Figure 8.1, one can visualize all wave frequencies that yield a single encounter frequency by drawing a horizontal line corresponding to the encounter frequency of interest and noting all intersections with the curves. Of course, only select relative wave directions are depicted in Figure 8.1, when in reality, many portions of the plot would yield continuous wave frequency functions ($\omega_1(\theta)$, $\omega_2(\theta)$, and $\omega_3(\theta)$) along these horizontal lines (discontinuous only where specific wave frequency functions yield non-real values). Given these three wave frequency functions corresponding to

ω_e, the fraction of energy attributable to each wave frequency function was desired. To calculate this quantity, it was first necessary to transform the 2D wave spectral densities from wave frequency to encounter frequency space, as shown in Equation 8.23.

$$S_e\big(\theta, \omega_i(\theta)\big) = \frac{g}{g - 2\big(\omega_i(\theta)\big)U\cos\big(\mu(\theta)\big)} S\big(\theta, \omega_i(\theta)\big) \tag{8.23}$$

Note that the transformation in Equation 8.23 continues to consider the wave frequencies appropriately as continuous functions of wave direction.

Next, it was necessary to identify the wave frequency function responsible for the wave candidate wave frequency considered by this series of filters. This identification of wave frequency function $\omega_j(\theta)$ is identified explicitly in Equation 8.24.

$$\omega_j(\theta) = \omega_i(\theta) \text{ s.t. } \omega_j\Big(\overline{\theta(\omega)}\Big) = \omega_i\Big(\overline{\theta(\omega)}\Big) = \omega \tag{8.24}$$

Given wave frequency function $\omega_j(\theta)$, the fraction of total energy at ω_e attributable to this specific wave frequency function could be calculated using Equation 8.25.

$$\begin{pmatrix}\text{Fraction of Energy at}\\ \omega_e \text{ due to } \omega_j(\theta)\end{pmatrix} = \frac{\int_0^{2\pi} S_e(\theta, \omega_j(\theta))\ d\theta}{\sum_{i=1}^{3}\int_0^{2\pi} S_e(\theta, \omega_i(\theta))\ d\theta} \tag{8.25}$$

Of course, at directions where $\omega_2(\theta)$ and $\omega_3(\theta)$ don't exist, spectral density values of zero were assigned for computation. Note that the actual implementation of Equation 8.25 followed the discrete computation presented in Equation 8.26.

$$\begin{pmatrix}\text{Fraction of Energy at}\\ \omega_e \text{ due to } \omega_j(\theta)\end{pmatrix} = \frac{\sum_{\Delta\theta} S_e\Big(\overline{\theta(\omega)} + \Delta\theta, \omega_j\Big(\overline{\theta(\omega)} + \Delta\theta\Big)\Big)}{\sum_{i=1}^{3}\sum_{\Delta\theta} S_e\Big(\overline{\theta(\omega)} + \Delta\theta, \omega_i\Big(\overline{\theta(\omega)} + \Delta\theta\Big)\Big)}, \tag{8.26}$$

$$\Delta\theta = -\pi, -\frac{179}{180}\pi,\ \ldots\ , \frac{179}{180}\pi$$

The fourth filter employed in this study is shown in Equation 8.27, which accepts a candidate wave frequency ω if greater than 99% of the energy at corresponding encounter frequency ω_e is attributable to wave frequency function $\omega_j(\theta)$.

$$\begin{pmatrix} \text{Fraction of Energy at} \\ \omega_e \text{ due to } \omega_j(\theta) \end{pmatrix} \geq 0.99 \tag{8.27}$$

Candidate wave frequencies that pass this filter yield relatively high confidence that energy in the encountered response spectrum can be traced to a small range of wave frequencies that surround the candidate wave frequency. The subtle variations in frequency across this small range are due to directional spreading (deemed acceptable via Filter 2) rather than the other two wave frequency functions. From a different perspective, this filter retains candidate wave frequencies in which energy attributed to the appropriate wave frequency function dominates energy attributable the other two wave frequency functions.

8.1.1.5 Filter 5: Encounter Frequency Directional Spreading Threshold

Although directional spreading was previously considered in Filter 2, given the high level of confidence required for reliable RAO calculation and the challenges associated with assigning response energy to one of the three wave frequency functions, the filter described here re-examines directional spreading, this time at encounter frequency ω_e (rather than a fixed wave frequency). The mean absolute directional difference across all three wave frequency functions corresponding to ω_e was calculated using Equation 8.28.

$$\overline{|\Delta\theta(\omega_e)|} = \frac{\sum_{i=1}^{3} \int_{-\pi}^{\pi} |\Delta\theta| S_e\left(\overline{\theta(\omega)} + \Delta\theta, \omega_i\left(\overline{\theta(\omega)} + \Delta\theta\right)\right) \, d(\Delta\theta)}{\sum_{i=1}^{3} \int_{-\pi}^{\pi} S_e\left(\overline{\theta(\omega)} + \Delta\theta, \omega_i\left(\overline{\theta(\omega)} + \Delta\theta\right)\right) \, d(\Delta\theta)} \tag{8.28}$$

As in previous filters, the above computation was actually performed using discrete values as shown in Equation 8.29.

$$\overline{|\Delta\theta(\omega_e)|} = \frac{\sum_{i=1}^{3}\sum_{\Delta\theta}|\Delta\theta|S_e\left(\overline{\theta(\omega)}+\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)}{\sum_{i=1}^{3}\sum_{\Delta\theta}S_e\left(\overline{\theta(\omega)}+\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)}, \quad \Delta\theta = -\pi, -\frac{179}{180}\pi, \ \ldots\ , \frac{179}{180}\pi \tag{8.29}$$

As performed for Filter 2, the present filter was based on a comparison to the cosine-squared spreading function. Therefore, as done for the wave spectrum in Equation 8.23, it was necessary to transform the cosine-squared spreading function to the encounter frequency space as if it were a spectral density function. This transformation is shown in Equation 8.30.

$$D_{e,cos^2}\left(\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right) = \frac{g}{g-2\left(\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)U\cos\left(\mu\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)} \quad D_{cos^2}(\Delta\theta) \tag{8.30}$$

Given this transformation, the mean absolute directional difference for the cosine-squared function at ω_e could be calculated using Equation 8.31.

$$\overline{|\Delta\theta_{cos^2}(\omega_e)|} = \frac{\sum_{i=1}^{3}\int_{-\pi}^{\pi}|\Delta\theta|D_{e,cos^2}\left(\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)\,d(\Delta\theta)}{\sum_{i=1}^{3}\int_{-\pi}^{\pi}D_{e,cos^2}\left(\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)\,d(\Delta\theta)} \tag{8.31}$$

Once again, this calculation was implemented discretely following Equation 8.32.

$$\overline{|\Delta\theta_{cos^2}(\omega_e)|} = \frac{\sum_{i=1}^{3}\sum_{\Delta\theta}|\Delta\theta|D_{e,cos^2}\left(\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)}{\sum_{i=1}^{3}\sum_{\Delta\theta}D_{e,cos^2}\left(\Delta\theta,\omega_i\left(\overline{\theta(\omega)}+\Delta\theta\right)\right)}, \quad \Delta\theta = -\pi, -\frac{179}{180}\pi, \ \ldots\ , \frac{179}{180}\pi \tag{8.32}$$

Note that unlike the Filter 2 calculations for spreading in the wave frequency point of view, Equations 8.31 and 8.32 did not yield a constant value because the existence of each wave frequency function was not guaranteed at a given wave direction, and the spreading function was transformed to the encounter frequency space based on speed and relative wave direction.

Similar to Filter 2, this filter directly compared mean absolute directional differences as shown in Equation 8.33.

$$\overline{|\Delta\theta(\omega_e)|} \leq \overline{|\Delta\theta_{cos^2}(\omega_e)|} \tag{8.33}$$

The filter in this step, Filter 5, confirmed that at the encounter frequency associated with candidate wave frequency ω, directional spreading across all three corresponding wave frequency functions was acceptable. Any candidate ω that passed these five filters was subject to one final filter based on the entire 2D wave spectrum.

8.1.1.6 Filter 6: Interpolated Wave Spectra Variance Threshold

The final wave data filter necessary in this section was based on consistency between the eight interpolated corners in latitude, longitude, and time as given in Equation 8.2. The purpose of this filter was to select points from the remaining pool of data that instilled relatively high confidence in the accuracy of the 2D spectral shape, which is extremely important for reliable RAO calculation. The assumption applied in

this stage is that consistency between the eight corners used in interpolation increases the likelihood that the interpolated 2D spectral shape is close to that experienced by the ship. A variance-based confidence metric, C, was employed, which is defined in Equation 8.34.

$$C^2 = \sum_{i=1}^{8} \sum_{j=1}^{24} \sum_{k=1}^{30} \left(S_{CDS,i}(\theta_j, \omega_k) - S_{CDS}(\theta_j, \omega_k) \right)^2 \tag{8.34}$$

This metric sums the squared differences between spectral densities at each of the eight corners and their respective interpolated values. The data filtering approach employed using this metric is demonstrated in Equation 8.35.

$$C \leq C_{Threshold} \tag{8.35}$$

Unlike the other five filters, this filter was based on the entire 2D wave spectrum rather focusing on a single wave frequency ω. As such, multiple data points that passed the first five filters independently could be removed simultaneously by this filter due to their shared 2D wave spectrum. Although this filter could have been applied first, it was saved as a final step such that $C_{Threshold}$ could be tuned to trim the pool of data to a manageable size for GPR model training. As such, the value selected for $C_{Threshold}$ in this study is discussed in Section 8.2.

8.1.2 Candidate Wave Frequencies Provided to Filters

All 10,034 AR Train windows discussed in Chapter VI with associated 2D CDS wave spectra were considered for measured RAO calculation in this study (excluding those with significant wave heights greater than two meters, as noted previously). As mentioned in the discussion of Filter 6, keeping the number of data points at a tractable level was important for training a GPR model without requiring approximation methods. In this study, the maximum amount of samples desired for GPR

training was between three and four thousand points. Therefore, it was not helpful to consider all 30 CDS wave frequencies as candidates for each of the 10,034 windows. Instead, for each of the 10,034 windows, local maxima were identified in the associated 1D wave spectrum calculated using Equation 8.4. These local maxima were selected as the candidate wave frequencies for a given window and subject to the six-step filtering procedure. This approach was deemed logical given that local energy peaks are often used to generate characteristic parameters for a given wave partition in multidirectional seas (e.g., spectral peak period). As such, the peaks represent a distinctive source of energy in the wave spectrum that is ideal for measured RAO calculation given that the peak passes all six filters.

8.1.3 Measured RAO Calculation

For any of the 10,034 AR Train windows that yielded one or more wave frequencies passing all six filters, Welch's method (*Welch*, 1967), implemented in the `SciPy` library (*Virtanen et al.*, 2020) of `Python`, was used to calculate the heave, pitch, and roll power spectral densities (a.k.a. response spectra), $S_{jM}(\omega_e)$, where "j" indicates the appropriate DOF. These spectra were functions of ω_e because they were experienced aboard the moving ship. Given these measured spectra and the 1D wave spectral density in encounter frequency space corresponding to an accepted candidate wave frequency, $S_e(\omega)$, measured heave, pitch, and roll RAO magnitudes at ω could be calculated as shown in Equation 8.36.

$$|RAO_{jM}(\omega)| = \sqrt{\frac{S_{jM}(\omega_e)}{S_e(\omega)}} \tag{8.36}$$

These RAO magnitudes were then used to train the GPR correction model as described in the following section.

8.1.4 GPR Model Construction and Data Standardization

Although the multiplicative correction factor approach of Chapter VII was deemed reasonably logical for that preliminary experiment, the limitations of the method given near-zero RAOs predicted using 2D strip theory (e.g., pitch in beam seas and roll in head seas) prompted a different formulation in this chapter. This study employed an additive correction approach, which was based on the difference between measured and 2D strip theory predicted RAO magnitudes, $\Delta|RAO_{jP}(\omega)|$, as outlined in Equation 8.37.

$$\Delta|RAO_{jP}(\omega)| = |RAO_{jM}(\omega)| - |RAO_{jP}(\omega)| \tag{8.37}$$

The difference calculated using Equation 8.37 for each accepted candidate wave frequency was used as a target value for GPR model training. Given this formulation, the output of the trained GPR model, $\Delta|RAO_{jGPR}(\omega)|$, could be used to calculate the GPR predicted RAO magnitude, $|RAO_{jGPR}(\omega)|$, as defined in Equation 8.38.

$$|RAO_{jGPR}(\omega)| = |RAO_{jP}(\omega)| + \Delta|RAO_{jGPR}(\omega)| \tag{8.38}$$

Without requiring the additional transformation defined in Equation 7.14 of Chapter VII, this formulation leveraged the zero-mean property of GPR such that uninformed regions of the GPR input space (meaning relatively far from any train data samples), reverted to the physics-based, 2D strip theory RAO magnitude, $|RAO_{jP}(\omega)|$. A single GPR model was constructed for each DOF (heave, pitch, and roll) in this study. Using the formulation presented here, this study exactly followed the GPR methodology presented in Section 7.2, including the use of the `Python` package `GPy` (*GPy*, since 2012).

The GPR models constructed for RAO correction using full-scale data in each

DOF employed five input variables. These five input variables are summarized in Table 8.1, where j indicates the DOF. Note that relative wave directions on (180°, 360°) were flipped to be on [0°, 180°] such that the GPR correction model leveraged port-starboard symmetry, which is consistent with the symmetry of the 2D strip theory RAOs. As shown in Table 8.1, in addition to the variables that RAOs are tradition-

Table 8.1: Input Variables for GPR Model Training.

Variable	Units	Description
U	m/s	Vessel Speed
$\mu\left(\overline{\theta(\omega)}\right)$	deg	Relative heading to the mean wave direction on [0°, 180°]
ω	rad/s	Wave Frequency
ω_e	rad/s	Encounter Frequency
$RAO_{jP}(\omega)\|$	units/m	2D strip theory predicted RAO magnitude

ally a function of (U, $\mu\left(\overline{\theta(\omega)}\right)$, and ω), the corresponding encounter frequency and 2D strip theory RAO magnitude were also provided as input the the GPR model. Although the underlying models are different, the inclusion of $|RAO_{jP}(\omega)|$ as an input variable resembles the implementations presented in *Weymouth and Yue* (2013) and *Weymouth* (2019). This incorporation of information from a physics-based model also resembles the use of PBMPs in NN and SLNN models presented in previous chapters and later discussion of this chapter.

The input data were standardized such that across the train dataset, each variable was scaled to have zero mean and a standard deviation equal to one. This standardization practice is commonly adopted for GPR and other machine learning models to promote numerical stability in model training. For the output variable, $|RAO_{jGPR}(\omega)|$, the mean was fixed to be zero (rather than the actual mean from the train data) and the standard deviation used in scaling was calculated using this assumed zero mean. This subtlety was important to preserve the aforementioned advantageous behavior in uninformed regions of the input space where GPR will output a value of zero.

8.1.5 Two-Stage Demonstration: GPR Corrected RAOs and the SLNN

After correcting the heave, pitch, and roll RAOs using their respective GPR models, these GPR RAOs were used in the initial twin framework to generate new PBMPs (referred to as GPR PBMPs) with each of the wave data sources from Chapter VI. These GPR PBMPs were then compared to the previous PBMPs presented in Chapter VI, which concluded the first stage of this model demonstration. Using these GPR PBMPs, the SLNN model construction, training, and testing procedure from Chapter VI was repeated for all three wave data sources, and the new results (labeled GPR+SLNN) were compared to those of Chapter VI. The use of the GPR PBMPs in the SLNN was the second and final stage of this demonstration, which was performed to investigate possible real-world application of these models in conjunction for improved predictive capabilities.

8.2 Results and Discussion

This section details the final filtered GPR dataset, the GPR-corrected RAOs, and the results of the two-stage model demonstration for heave, pitch, and roll prediction.

8.2.1 Filtered Dataset

As mentioned in Section 8.1.1, the variance-based confidence metric threshold $C_{Threshold}$ of Filter 6 was tunable to trim the pool of data to a manageable size for GPR model training. Figure 8.2 shows the cumulative mass function (CMF) of C for the approximately 3,000 samples remaining after Filter 5. Although 3,000 points would not have yielded unreasonable computation time for GPR model training, there is a sharp bend in the CMF where C increases rapidly above approximately 2,500 samples. As such, $C_{Threshold}$, was set at a value of 0.01 as indicated by the black vertical line in Figure 8.2, which yielded 2,510 samples. Recall that the purpose of

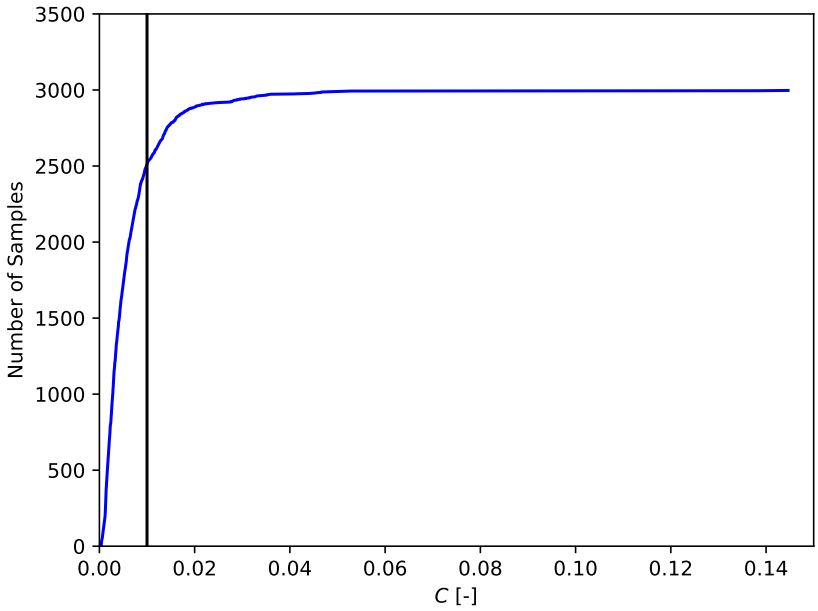

Figure 8.2: Cumulative mass function (CMF) of Filter 6's confidence metric, C, for the samples remaining after Filter 5. The vertical line indicates the selected filter threshold, $C_{Threshold}$.

Filter 6 was to select wave data that instilled relatively high confidence in the accuracy of the 2D spectral shape based on consistency between the eight interpolation corners. Note that given more candidate wave frequencies considered for filtering (i.e., more than just local maxima of 1D wave spectra), this threshold would be necessary for true trimming of the dataset for feasible GPR training rather than just removing points with the largest C values.

To put Filter 6's trimming of approximately 500 samples in context, Table 8.2 summarizes the six filters and the percentage of total candidate samples remaining after each filter. As shown in Table 8.2, Filters 1, 2, 4, and 6 each trimmed a substantial amount of samples. However, it is important to note that the percentage of candidates remaining after a given filter was also dependent on the number of candidates provided to it by prior filters. For instance, consider Filters 2 and 5, which both addressed directional spreading. Filter 5's removal of a small number of

Table 8.2: Summary of the six wave data filters and the percentage of total candidates that passed each filter.

Filter Number	Name	Percentage of Candidates Remaining After Filter
1	1D Spectral Density Threshold	89.4%
2	Directional Spreading Threshold	51.9%
3	Encounter Frequency Nyquist Threshold	51.8%
4	Encounter Frequency Energy Threshold	29.2%
5	Encounter Frequency Directional Spreading Threshold	28.0%
6	Interpolated Wave Spectra Variance Threshold	23.5%

points relative to the number trimmed by Filter 2 does not mean that points failing to pass Filter 2 would not have also failed to pass Filter 5. In short, while the order of the filters did not matter in terms of the resultant dataset, it did matter for the values provided in Table 8.2; therefore, the provided percentages do not directly speak to the importance of each filter, but rather serve as a demonstration of the filtering approach.

Figure 8.3 shows the relative wave direction (on [0°, 180°]) and speed of the 2,510 samples used in GPR model training. Unsurprisingly, Figure 8.3 shows higher point densities closer to zero speed and the vessel's design speed of 12 knots than at intermediate speeds. For following seas (approaching 0°) there is far greater sample density seen for lower speeds than at higher speeds. Despite the filtering approach's efforts and successes to retain data for relative directions less than 90°, there are still clearly limitations in terms of data suitable for RAO correction as the seas approach 0°. Nevertheless, there is still a substantial number of samples across the range of speeds that fall below 90° where multiple wave frequencies correspond to the same encounter frequency for non-zero forward speeds. As such, the developed filtering approach had notable success selecting data points from full-scale data that were not necessarily usable for RAO correction at first glance.

Figures 8.4 and 8.5 show relative wave direction and speed versus wave frequency,

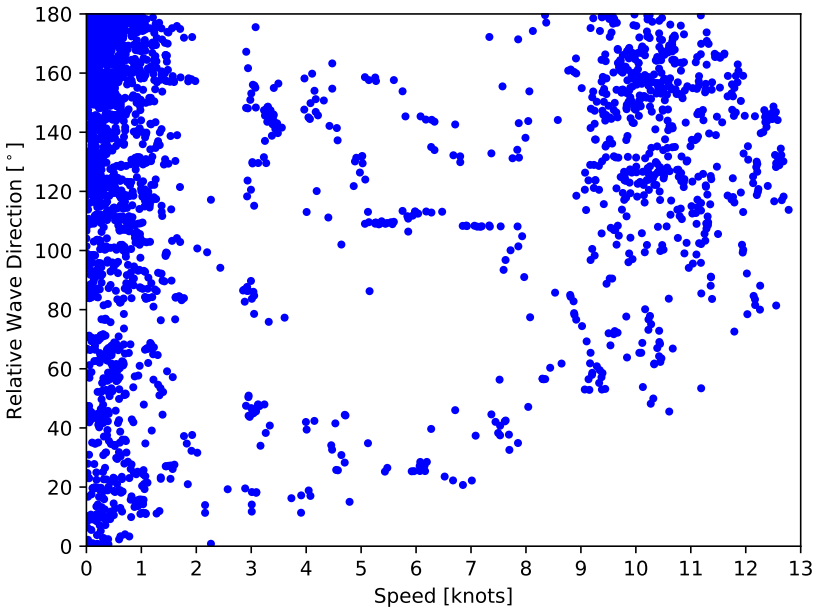

Figure 8.3: Relative wave direction versus speed for the 2,510 samples that passed all six filters and were used to train the GPR RAO correction model.

respectively, for the 2,510 samples. The vertical lines indicate the discrete wave frequencies of the CDS ERA5 2D wave spectra defined in Equation 8.1. These indicated wave frequencies could have been under consideration by the six filters given a corresponding local maximum in the 1D spectrum for any particular AR Train window.

Figures 8.4 and 8.5 were included to demonstrate the spread of data across the relatively large input space for the GPR model (with respect to three of the five input variables). Additionally, these figures show that no wave frequencies above 2 rad/s passed the six filters. The GPR models trained for heave, pitch, and roll RAO correction using these samples are detailed in the following section.

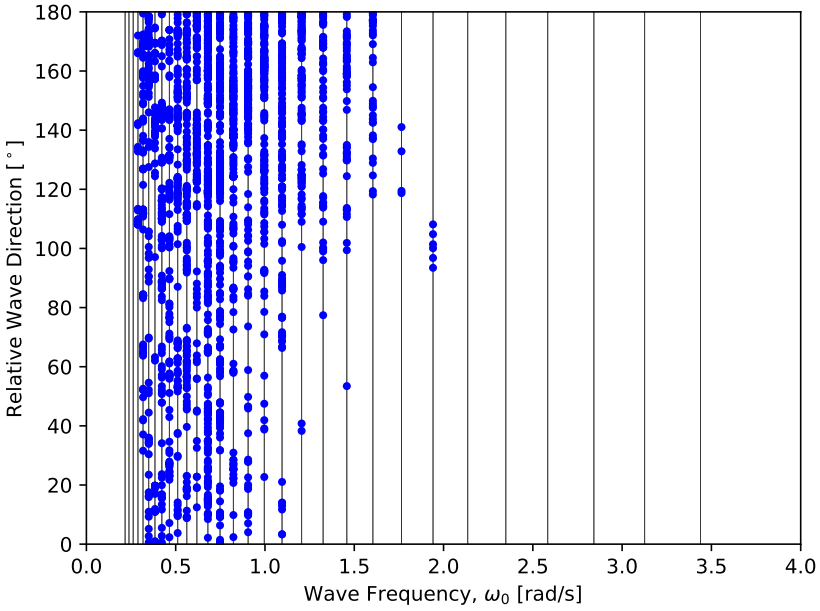

Figure 8.4: Relative wave direction versus wave frequency for the 2,510 samples that passed all six filters and were used to train the GPR RAO correction model.

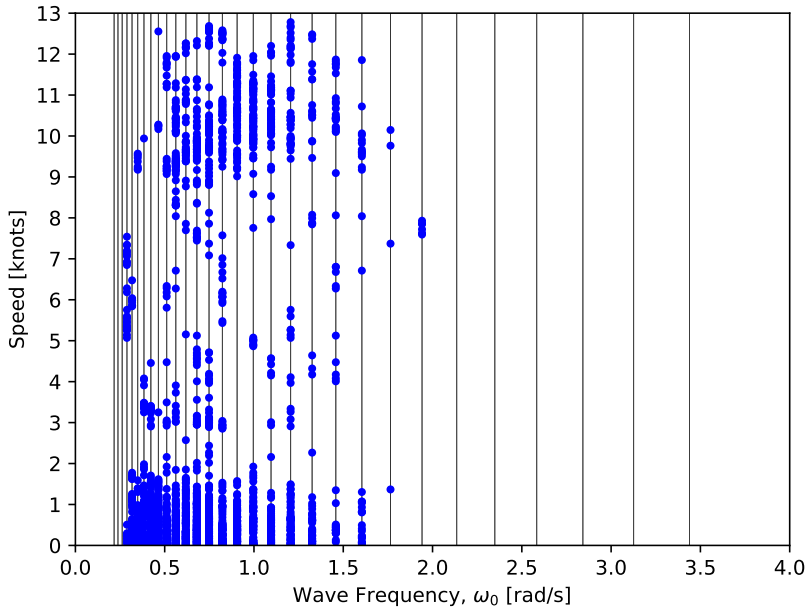

Figure 8.5: Vessel speed versus wave frequency for the 2,510 samples that passed all six filters and were used to train the GPR RAO correction model.

8.2.2 GPR RAO Correction

Unlike the GPR model trained in Chapter VII, the training process in this study consistently yielded identical hyperparameters for this model formulation and train dataset. The radial basis function (RBF) lengthscales (based on standardized input variable values) for the five input variables and each GPR model (heave, pitch, and roll) are summarized in Table 8.3.

Table 8.3: Trained GPR model RBF lengthscales (based on standardized values).

	GPR RBF Lengthscales		
Variable	Heave	Pitch	Roll
U	0.201	0.285	14.238
$\mu\left(\overline{\theta(\omega)}\right)$	0.167	0.635	7.254
ω	0.120	0.212	0.310
ω_e	103.052	14.179	10.760
$\lvert RAO_{jP}(\omega)\rvert$	8,848.556	0.417	2.554

As explained by *Rasmussen and Williams* (2006), as the lengthscale becomes longer for a given input variable, the covariance becomes increasingly independent of that variable. Therefore, shorter lengthscales indicate greater importance for GPR prediction. Correspondingly, variables with exceptionally long lengthscales could be omitted with little to no consequence to the model, which is the theory behind the automatic relevance determination (ARD) mentioned in Section 7.2.

Although data for each input variable were standardized, note that the lengthscales are still dependent on the distribution of each input variable in the train dataset, and unstandardized versions of the lengthscales in Table 8.3 would depend on each variable's units; therefore, direct comparisons between lengthscales of different variables should be made carefully. That said, the relatively large lengthscales for the encounter frequency and 2D strip theory RAO magnitude variables in the heave GPR model indicate that they were of lesser importance than the other three input variables. For pitch, encounter frequency was of lesser importance based on the

lengthscale. It's also interesting that the 2D strip theory pitch RAO magnitudes had a lengthscale indicative of importance to the model. The relatively high quality of the *Neil Armstrong*'s pitch RAOs discussed in previous chapters may have spurred their usefulness for the GPR model. For roll, wave frequency had the shortest lengthscale, followed by 2D strip theory RAO magnitude, indicating their relative importance to the model. Although the roll RAOs were very low quality as discussed in previous chapters, note that these poor RAOs could still provide relationships between train data samples that benefit the model. Furthermore, the fact that speed and relative wave direction had longer lengthscales for roll does not mean that the ship's roll RAOs are not a function of these variables. On the contrary, these longer lengthscales just mean that these two variables were not particularly relevant for predicting the difference between the measured and 2D strip theory RAOs (Equation 8.37).

Figure 8.6 shows the uncorrected (SHIPMO) and GPR-corrected heave, pitch, and roll RAOs for five relative headings at zero speed. The 95% confidence interval (CI) resulting from the GPR model's output variance is also included, as well as the measured RAOs used in model training within 1 knots and 5° of the plotted RAOs. The GPR RAOs for all three DOFs are notably different from the uncorrected RAOs in several instances. Unsurprisingly, these differences are most pronounced for roll, where the uncorrected RAOs were of relatively low quality. The CIs are logically narrower near train data points. Note the seemingly sinusoidal fluctuation for the roll CI at select wave frequencies. These fluctuations are attributed to the relatively short wave frequency lengthscale and contrastingly long lengthscales for speed, relative direction, and encounter frequency, causing rapid changes in the wave frequency direction. For instance, the GPR RAO for roll at 90° still contains a spike corresponding to the uncorrected RAO's peak due to this short lengthscale despite train data samples that are in the vicinity. Although heave and pitch also had relatively short wave frequency lengthscales, they also have comparably short

lengthscales in other input dimensions.

In Figure 8.6, it should be noted that there are a few instances where the RAOs (roll, in particular) briefly dropped to negative values. Although these instances were infrequent across the full range of speeds and headings, and the corresponding negative RAO values were relatively insignificant, this behavior is non-physical; as such, it is recommended for future studies that any GPR-corrected RAO values less than zero be set to zero. Also, because this behavior is non-physical, it follows that none of the measured RAOs used in training were calculated to be negative, but the employed difference formulation of the GPR target variable (Equation 8.37) and the learned hyperparameters made negative predicted values possible.

As shown in Figure 8.6, there was an abundance of data available for most relative directions and speeds near 0 knots. As indicated in Figures 8.3 and 8.5, the amount of data available at the design speed of 12 knots and other intermediate speeds was not as substantial. Figures 8.7 and 8.8 reflect this lesser amount of data in RAO plots identical to Figure 8.6 at speeds of 9 knots and 12 knots, respectively. In both figures, there are notable differences between the uncorrected and GPR-corrected RAOs in several instances. Once again, these differences were most notable for roll, which held across all five included directions. This consistency across directions despite a lack of train data in stern-quartering (45°) and following (0°) seas is attributed to the relative importance of the strip theory RAO input variable shown by the corresponding lengthscale in Table 8.3. In contrast, for heave and pitch in stern-quartering and following seas, the GPR-corrected RAOs were nearly identical to the uncorrected, physics-based RAOs owing to a lack of relevant train data and the lengthscale-implied importance of speed and relative wave direction. Although relevant data would be preferred, the models' return to the uncorrected RAOs rather than yielding erroneous results is extremely desirable. As mentioned previously, this advantageous behavior was made possible by GPR's zero-mean property and the difference formulation de-

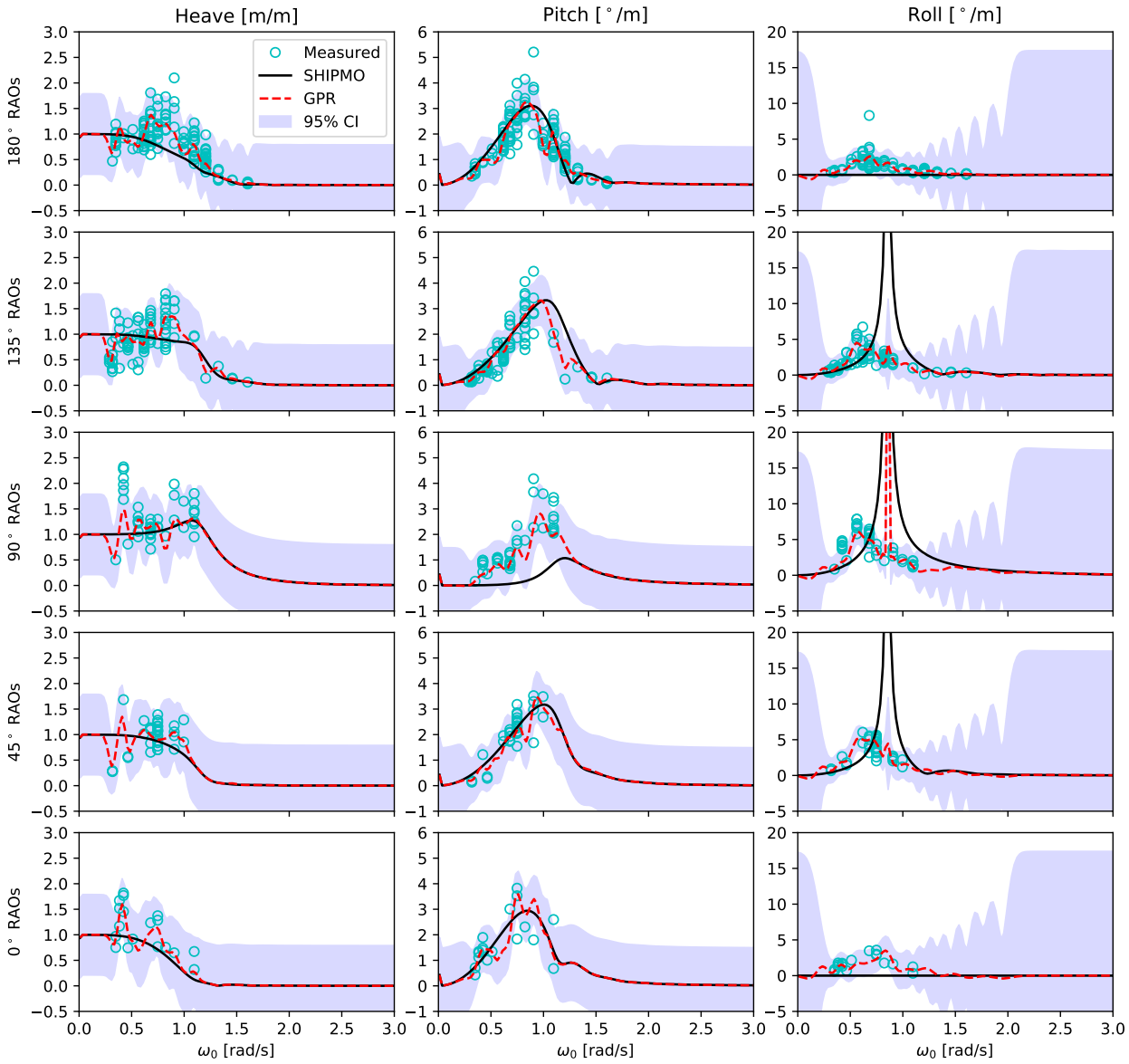

Figure 8.6: The uncorrected (SHIPMO) and GPR-corrected heave, pitch, and roll RAOs for five relative headings at 0 knots. The 95% confidence interval resulting from the GPR model's output variance is also included, as well as the measured RAOs used in model training within 1 knots and 5° of the plotted RAOs.

fined in Equation 8.37. These GPR-corrected heave, pitch, and roll RAOs, which employed filter-selected data from the AR Train dataset, were then used to generate AR Train, AR Test, and SR Test PBMPs for all three wave data sources of Chapter VI, with results presented in the following section.

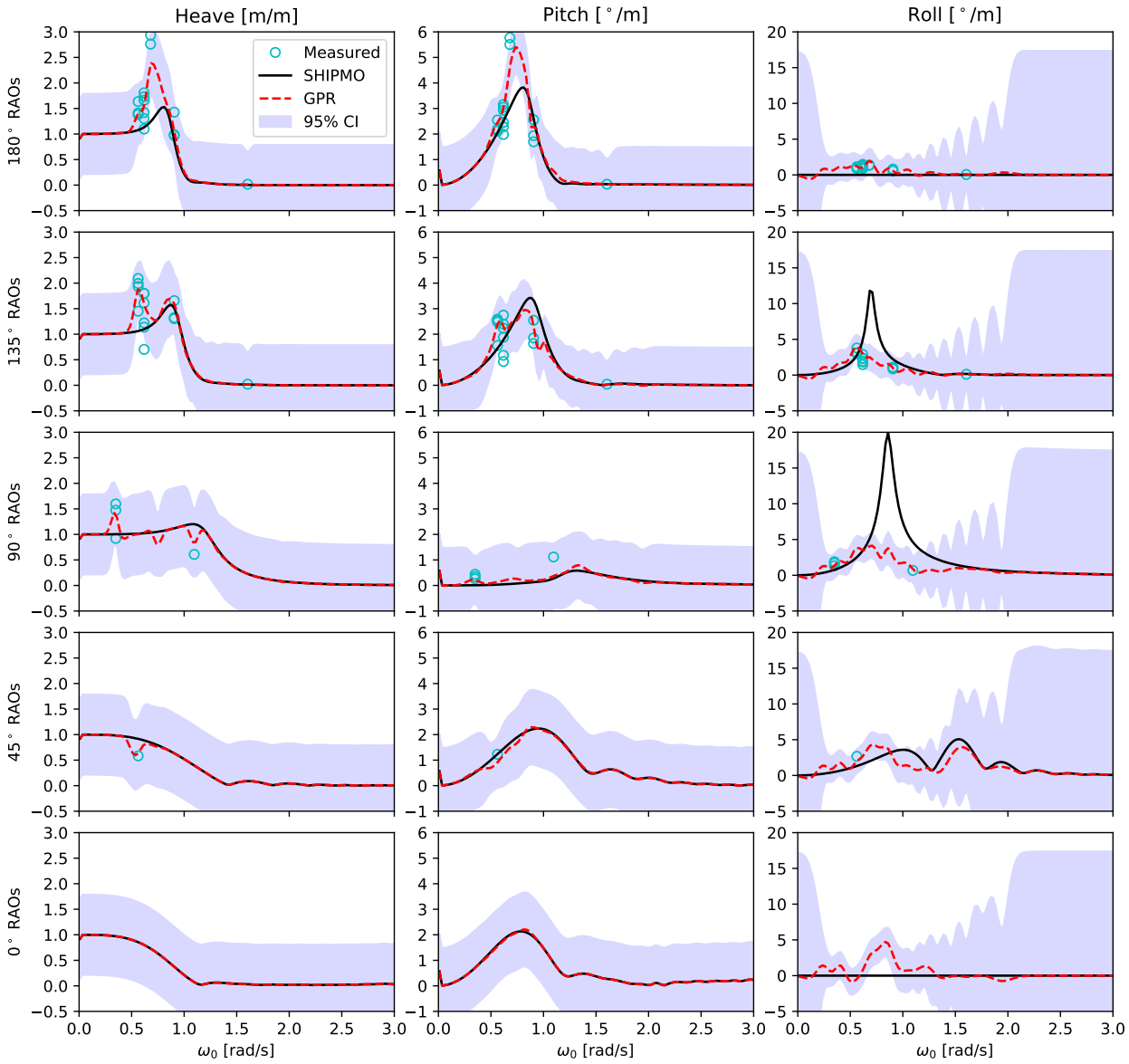

Figure 8.7: The uncorrected (SHIPMO) and GPR-corrected heave, pitch, and roll RAOs for five relative headings at 9 knots. The 95% confidence interval resulting from the GPR model's output variance is also included, as well as the measured RAOs used in model training within 1 knots and 5° of the plotted RAOs.

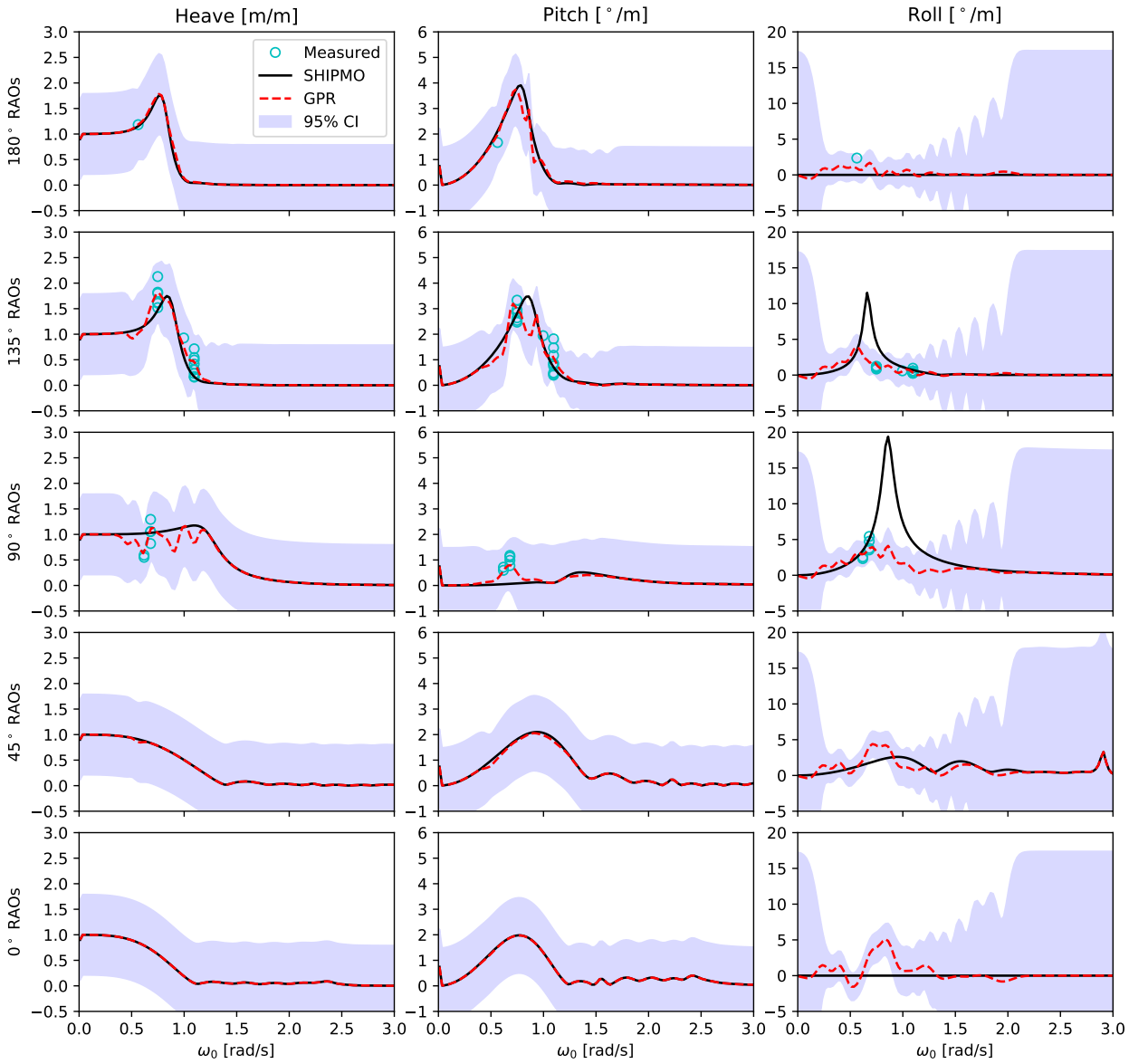

Figure 8.8: The uncorrected (SHIPMO) and GPR-corrected heave, pitch, and roll RAOs for five relative headings at 12 knots. The 95% confidence interval resulting from the GPR model's output variance is also included, as well as the measured RAOs used in model training within 1 knots and 5° of the plotted RAOs.

8.2.3 GPR PBMPs and their use in the SLNN

While the GPR RAOs presented in Figures 8.6, 8.7 and 8.8 seem logical based on the measured RAOs, comparisons of PBMPs generated using these RAOs to those generated using the uncorrected RAOs were necessary to determine the efficacy of the GPR correction models. This section simultaneously presents this PBMP comparison, which concluded the first demonstration stage of this study, and a performance comparison between SLNNs trained and tested using these GPR PBMPs and the standard PBMPs, which comprised the second stage of this demonstration. For both comparisons, the PBMP and SLNN results from Chapter VI are reproduced in all remaining tables and figures of this chapter.

For each wave data source, Table 8.4 summarizes the 10-fold CV-selected SLNN models from Chapter VI and those trained using the GPR PBMPs. It is interesting to note that for all three wave data sources, the GPR+SLNN models yielded greater average CV losses than their counterparts with uncorrected PBMPs.

Table 8.4: Results of the 10-fold CV model selection processes for each data-driven model. These selections were performed exclusively using the data in AR Train. Rows with "GPR+" were trained with PBMPs generated using GPR-corrected RAOs, while the other rows were presented in Chapter VI.

Wave Data Source	Model	Average CV Loss	Std. Dev. of CV Loss	λ	HL Architecture
CMEMS	SLNN	0.001675	0.001176	0.000130	[100, 100]
	GPR+SLNN	0.001703	0.001255	0.000100	[100, 100]
CDS	SLNN	0.001905	0.001302	0.000530	[60, 60]
	GPR+SLNN	0.001957	0.001320	0.000600	[120, 120]
NWW3	SLNN	0.001942	0.001097	0.000055	[100, 100, 100]
	GPR+SLNN	0.002032	0.001286	0.000070	[100, 100]

Table 8.5 shows the AR Train heave, pitch, roll, and total variance MSE values for each of the wave data sources. Although train data is not a strong indicator of real-world performance for the SLNN models, comparison of the PBMPs presented

in Table 8.5 is a worthwhile analysis. Recall that the measured RAOs used in GPR model training resulted from a filter-selected subset of the AR Train dataset, so analyzing the GPR PBMPs' performance with AR Train is not a fully independent evaluation. Nevertheless, a large majority of the 10,034 samples in AR Train had no representation in the GPR model's final 2,510 measured RAO values. As such, it is interesting to note that for heave and roll, the GPR PBMPs yielded lower variance MSE values than the standard PBMPs for all three wave data sources, demonstrating the efficacy of the GPR RAO correction approach. On the contrary, for pitch, the GPR PBMPs had greater variance MSE values than the standard PBMPs for all three wave data sources, implying that the GPR correction was slightly detrimental in this DOF. As with the results in Chapter VI, it is likely that the motion variance MSE metric does not tell the full picture of model performance. Therefore, it is important to further this discussion of model performance using the motion amplitude RMSPE and RMSE metrics in later discussion.

Table 8.6 shows the SLNN direct route slopes pre- and post-training. As mentioned in Chapter VI, although the conclusions that can be drawn from these slopes are limited given that the FCN of each SLNN could also account for linear relationships between the PBMPs and model output, the slopes assigned to each DOF's PBMPs are intriguing. In Table 8.6, the most interesting features are the greater roll slopes seen for the GPR+SLNN models than their SLNN counterparts. This increase in slopes both pre- and post-training implies that greater weight was placed on the GPR-corrected PBMPs than the uncorrected PBMPs, which further suggests that the roll GPR model improved the RAOs and resultant GPR PBMPs.

Table 8.5: AR Train performance of the data-driven models and their underlying uncorrected or GPR-corrected PBMPs (GPR PBMPs). As indicated, the MSE values are those of the normalized motion variances, which were used in model training, rather than normalized motion amplitudes. Rows with "GPR+" were trained using GPR PBMPs, while the other rows were presented in Chapter VI.

Wave Data Source	Model	Heave Var. MSE [-]	Pitch Var. MSE [-]	Roll Var. MSE [-]	Total Var. MSE [-]
CMEMS	PBMPs	0.002541	0.002072	0.069400	0.024671
	GPR PBMPs	0.001752	0.002290	0.003103	0.002382
	SLNN	0.001089	0.001713	0.001184	0.001330
	GPR+SLNN	0.001116	0.001603	0.001202	0.001308
CDS	PBMPs	0.002263	0.002190	0.117271	0.040575
	GPR PBMPs	0.001702	0.002370	0.003251	0.002441
	SLNN	0.001235	0.002099	0.001403	0.001581
	GPR+SLNN	0.001335	0.002174	0.001395	0.001636
NWW3	PBMPs	0.002779	0.002594	0.089732	0.031702
	GPR PBMPs	0.001994	0.003039	0.003068	0.002700
	SLNN	0.001179	0.001866	0.001136	0.001396
	GPR+SLNN	0.001220	0.001864	0.001199	0.001430

Table 8.6: SLNN direct-route slopes (SL Slopes) pre- and post-training. Rows with "GPR+" used GPR PBMPs, while the other rows were presented in Chapter VI.

Wave Data Source	Model	**Heave** Pre-	**Heave** Post-	**Pitch** Pre-	**Pitch** Post-	**Roll** Pre-	**Roll** Post-
CMEMS	SLNN	1.577	1.016	1.061	0.680	0.234	0.105
	GPR+SLNN	1.296	0.859	1.093	0.758	1.052	0.505
CDS	SLNN	1.460	1.388	0.966	0.880	0.169	0.049
	GPR+SLNN	1.211	0.960	0.998	0.815	0.830	0.492
NWW3	SLNN	1.650	1.060	0.996	0.602	0.215	0.037
	GPR+SLNN	1.332	0.841	1.034	0.720	1.025	0.469

Table 8.7 shows the AR Test variance MSE values for each model and the underlying PBMPs or GPR PBMPs. As seen for the AR Train results, for all three wave data sources, the GPR PBMPs yielded significantly lower roll variance MSE values than the standard PBMPs. For heave, this improvement was seen for CMEMS and CDS wave data, but not NWW3. For pitch, the standard PBMPs notably outperformed the GPR PBMPs for all three wave data sources, suggesting that the GPR RAO correction was detrimental. While the results suggest that the GPR RAO correction did not consistently improve upon heave and pitch predictive capabilities in terms of variance MSE reduction, further motion amplitude RMSPE and RMSE analysis was warranted and included in later discussion.

Comparing the SLNN results to the GPR+SLNN results for AR Test in Table 8.7, the heave and pitch variance MSE values were typically greater for the GPR+SLNN, suggesting that the GPR-corrected RAOs hurt SLNN model performance. For roll, there was a very slight reduction in variance MSE values across all wave data sources for the GPR+SLNN model. It's also interesting to note that for each DOF and wave data source in Table 8.7, the best-performing SLNN always outperformed the best-performing PBMPs, further supporting the developed SLNN structure's efficacy. Again, performance evaluations using model amplitude RMSPE and RMSE were necessary for a more definitive picture of model performances.

Table 8.8 shows the motion variance MSE values for each model and the underlying PBMPs or GPR PBMPs of SR Test. For heave, the GPR PBMPs yielded greater variance MSE for all three wave data sources than the standard PBMPs. On the contrary, for pitch, the GPR PBMPs produced lower variance MSE values across wave data sources, which contrasts the results for AR Train and AR Test. For roll, the GPR PBMPs significantly reduced the variance MSE values relative to the standard PBMPs, suggesting that the GPR RAO correction was very valuable for roll.

Comparing SLNN and GPR+SLNN performances in Table 8.8, for heave and roll,

Table 8.7: AR Test performance of the data-driven models and their underlying uncorrected or GPR-corrected PBMPs (GPR PBMPs). As indicated, the MSE values are those of the normalized motion variances, which were used in model training, rather than normalized motion amplitudes. Rows with "GPR+" were trained using GPR PBMPs, while the other rows were presented in Chapter VI.

Wave Data Source	Model	Heave Var. MSE [-]	Pitch Var. MSE [-]	Roll Var. MSE [-]	Total Var. MSE [-]
CMEMS	PBMPs	0.003651	0.002268	0.078923	0.028280
	GPR PBMPs	0.002955	0.005861	0.006959	0.005258
	SLNN	0.001160	0.001566	0.003705	0.002146
	GPR+SLNN	0.001390	0.002099	0.003702	0.002400
CDS	PBMPs	0.003860	0.001917	0.082919	0.029565
	GPR PBMPs	0.002665	0.004113	0.007825	0.004868
	SLNN	0.001613	0.001886	0.004886	0.002804
	GPR+SLNN	0.001920	0.002928	0.004589	0.003156
NWW3	PBMPs	0.004136	0.006239	0.139208	0.049861
	GPR PBMPs	0.004710	0.020487	0.007874	0.011024
	SLNN	0.002159	0.003127	0.004074	0.003091
	GPR+SLNN	0.002500	0.005145	0.003919	0.003843

the GPR+SLNN yielded lower variance MSE values across all three wave data sources. For pitch, the GPR+SLNN produced lower variance MSE values for CMEMS and CDS, but not NWW3. Note that unlike the results for AR Test, the best-performing SLNN did not always outperform the best-performing PBMPs for SR Test. This may suggest that the physics-based approaches generalize better between sister vessels than the SLNN in some cases. As mentioned previously, the preliminary conclusions drawn from AR Train, AR Test, and SR Test using motion variance MSE values warranted further analyses using the motion amplitude RMSPE and RMSE metrics, which are provided in the following paragraphs.

Table 8.8: SR Test performance of the data-driven models and their underlying uncorrected or GPR-corrected PBMPs (GPR PBMPs). As indicated, the MSE values are those of the normalized motion variances, which were used in model training, rather than normalized motion amplitudes. Rows with "GPR+" were trained using GPR PBMPs, while the other rows were presented in Chapter VI.

Wave Data Source	Model	Heave Var. MSE [-]	Pitch Var. MSE [-]	Roll Var. MSE [-]	Total Var. MSE [-]
CMEMS	PBMPs	0.000363	0.000323	0.005781	0.002156
	GPR PBMPs	0.000412	0.000290	0.000822	0.000508
	SLNN	0.000555	0.000357	0.000595	0.000503
	GPR+SLNN	0.000373	0.000276	0.000520	0.000390
CDS	PBMPs	0.000339	0.000232	0.004925	0.001832
	GPR PBMPs	0.000354	0.000222	0.000663	0.000413
	SLNN	0.000448	0.000231	0.000587	0.000422
	GPR+SLNN	0.000299	0.000225	0.000519	0.000348
NWW3	PBMPs	0.000341	0.000458	0.012194	0.004331
	GPR PBMPs	0.000439	0.000410	0.000748	0.000532
	SLNN	0.000566	0.000359	0.000664	0.000530
	GPR+SLNN	0.000528	0.000488	0.000510	0.000509

Figure 8.9 shows the GPR PBMPs and GPR+SLNN motion amplitude RMSPE results overlaid on those from Figure 6.2 for the CMEMS wave data source. Recall that the large RMSPE values in the lowest motion amplitude bins were due primarily to low measured motion amplitudes rather than large error values, as shown in the corresponding RMSE plots of Figure 8.10. In Figures 8.9 and 8.10, for heave and roll, the GPR PBMPs typically outperformed the standard PBMPs for AR Train and AR Test. The pitch GPR PBMPs of AR Test were very similar to the standard PBMPs for several bins, but yielded greater RMSPE and RMSE for several higher motion amplitude bins, showing a breakdown in performance at higher amplitudes. For heave and pitch of the SR Test dataset, the GPR PBMPs and standard PBMPs were very similar in most bins. With the exception of the two greatest bins in SR Test, which had very few samples, the GPR PBMPs outperformed the standard PBMPs for roll, suggesting that the GPR correction approach was effective.

Regarding the AR Test and SR Test SLNN and GPR+SLNN curves in Figures 8.9 and 8.10, both outperformed the PBMPs across wave data sources and DOFs in most bins. This result further emphasizes the predictive power of the developed SLNN model structure. Comparing the two SLNNs, both produced similar results in all cases, with select instances where the standard SLNN outperformed the GPR+SLNN. As such, it is not evident that the use of GPR PBMPs within the SLNN structure (i.e., GPR+SLNN) had any benefit for the CMEMS wave data source.

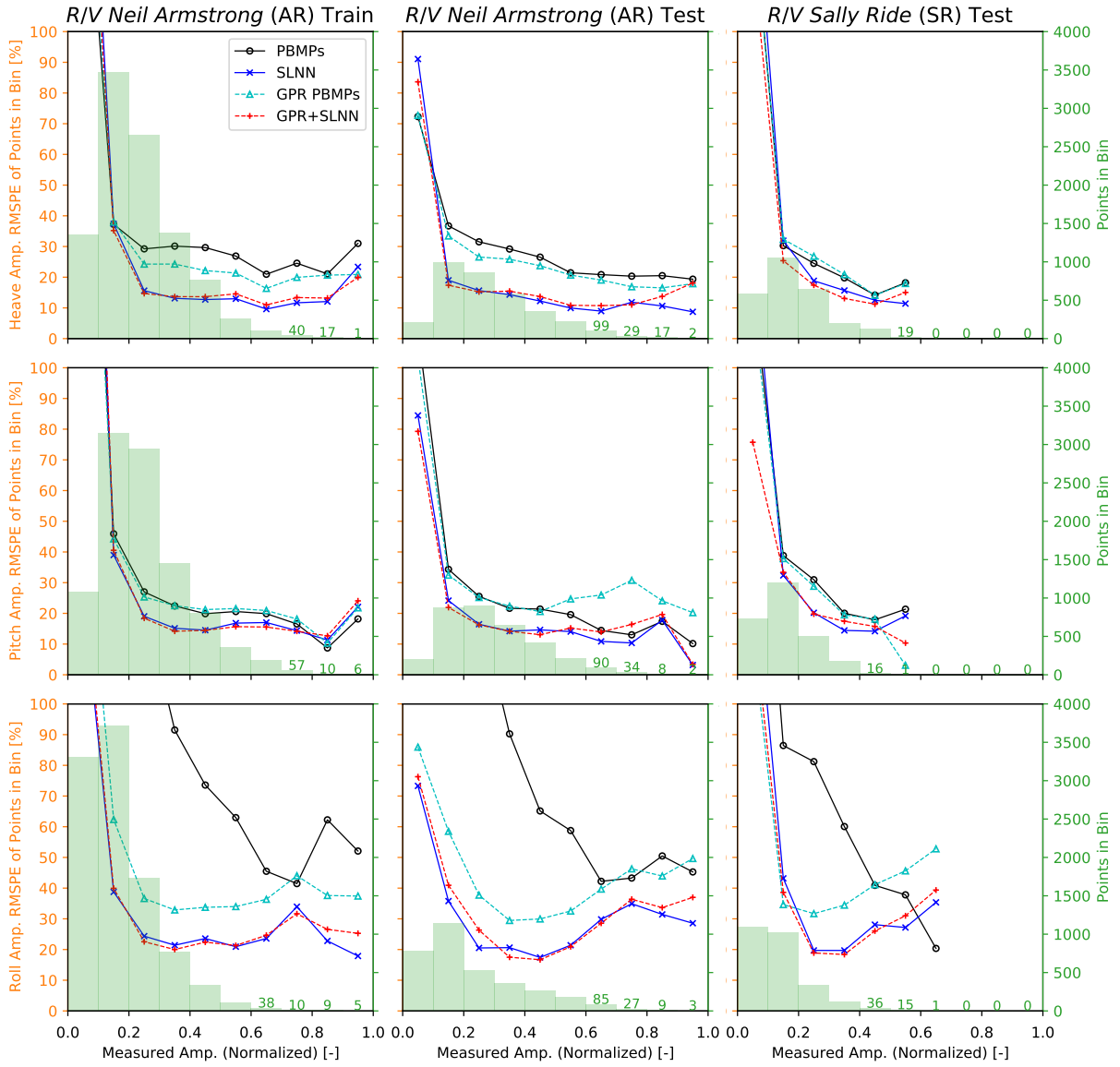

Figure 8.9: Plots of heave, pitch, and roll RMSPE as a function of their respective normalized motion amplitude measurements for the CMEMS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSPE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

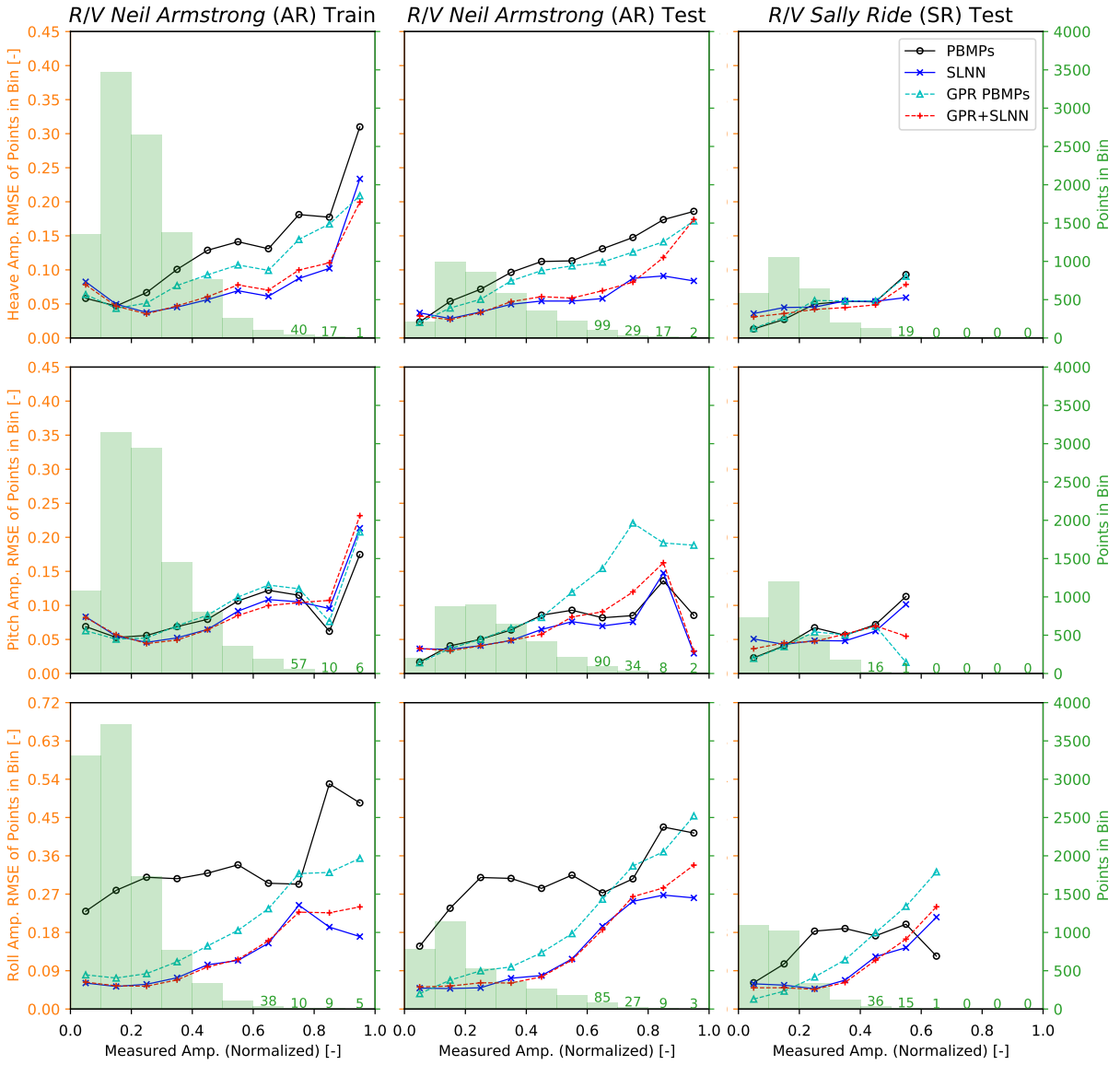

Figure 8.10: Plots of heave, pitch, and roll RMSE as a function of their respective normalized motion amplitude measurements for the CMEMS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

Figures 8.11 and 8.12 show the motion amplitude RMSPE and RMSE values for the CDS wave data source corresponding to Figures 6.4 and 6.5. For AR Test heave, beyond the lowest amplitude bins, the GPR PBMPs outperformed the standard PBMPs. For SR Test heave and pitch, both PBMP curves were very similar, excluding the highest pitch bin, which had just one sample. For AR Test pitch, the PBMP curves were very similar until the higher amplitude bins, where the standard PBMPs outperformed the GPR PBMPs in most cases. This poorer performance for the GPR PBMPs was also seen for the CMEMS wave data source, showing that the GPR correction was somewhat detrimental for pitch at these higher amplitudes. For roll in both AR Test and SR Test of Figures 8.11 and 8.12, the GPR PBMPs were dramatically better than the standard PBMPs in intermediate amplitude bins, but showed poorer performance than the standard PBMPs in the highest amplitude bins. This behavior was also witnessed for roll in the AR Train dataset, showing consistency in the GPR PBMPs' performance. This behavior could be due to the inherent nonlinear complexities of roll motion. As such, it is entirely possible that these GPR corrected RAOs are not problematic, but rather they break down as linear assumptions become less valid. Some of the relatively poor pitch performances at higher amplitudes seen for CMEMS and CDS may also be due to this breakdown in linear theory, although the standard pitch PBMPs seem to continue with relatively strong performance.

Comparing the SLNN curves in Figures 8.11 and 8.12, both models performed very similar to each other across all DOFs and datasets for CDS, showing no distinct advantage to using GPR PBMPs over standard PBMPs. Nevertheless, both SLNNs outperformed the PBMPs in most bins across DOFs, reinforcing the predictive power of the SLNN structure.

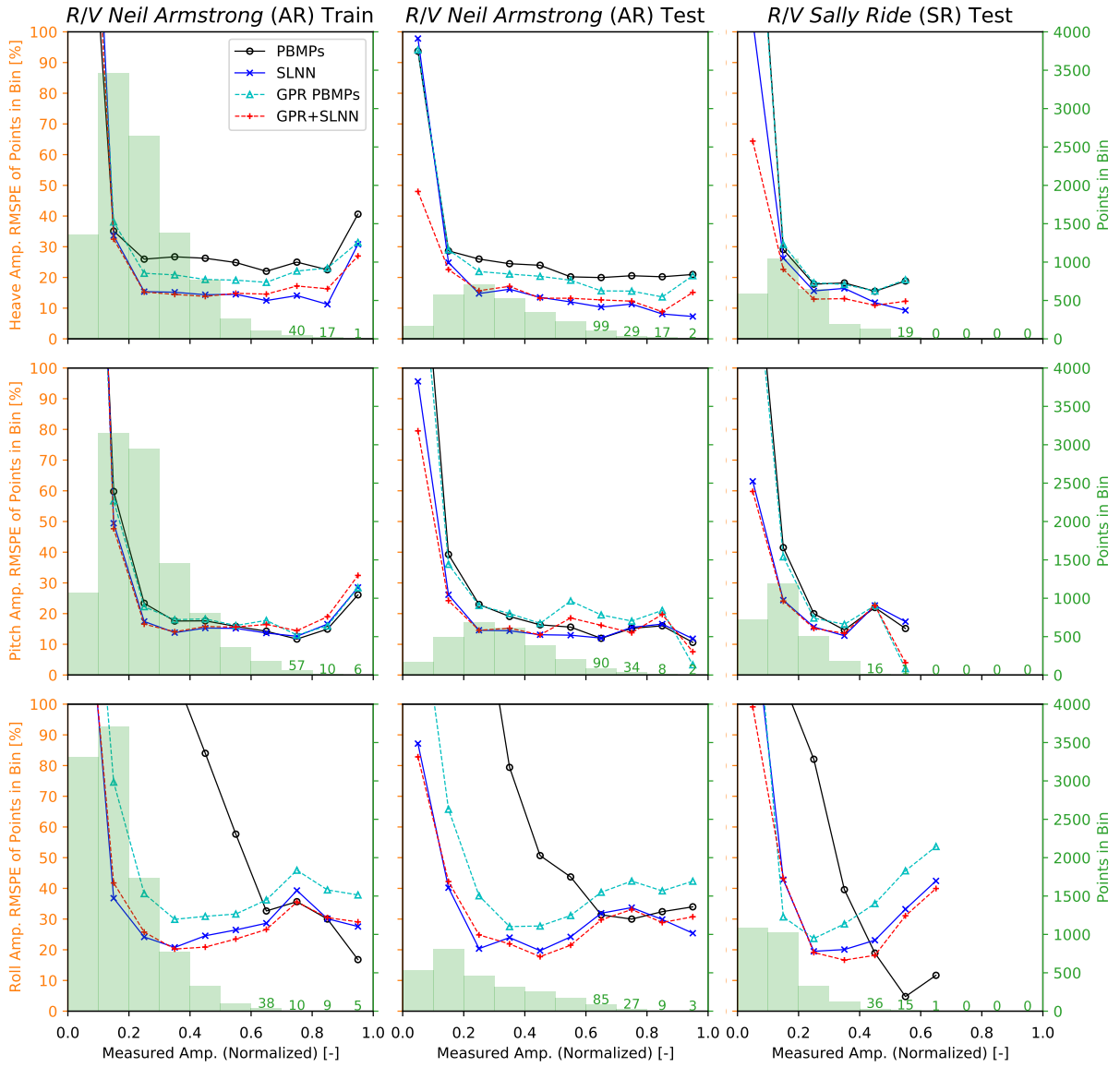

Figure 8.11: Plots of heave, pitch, and roll RMSPE as a function of their respective normalized motion amplitude measurements for the CDS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSPE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

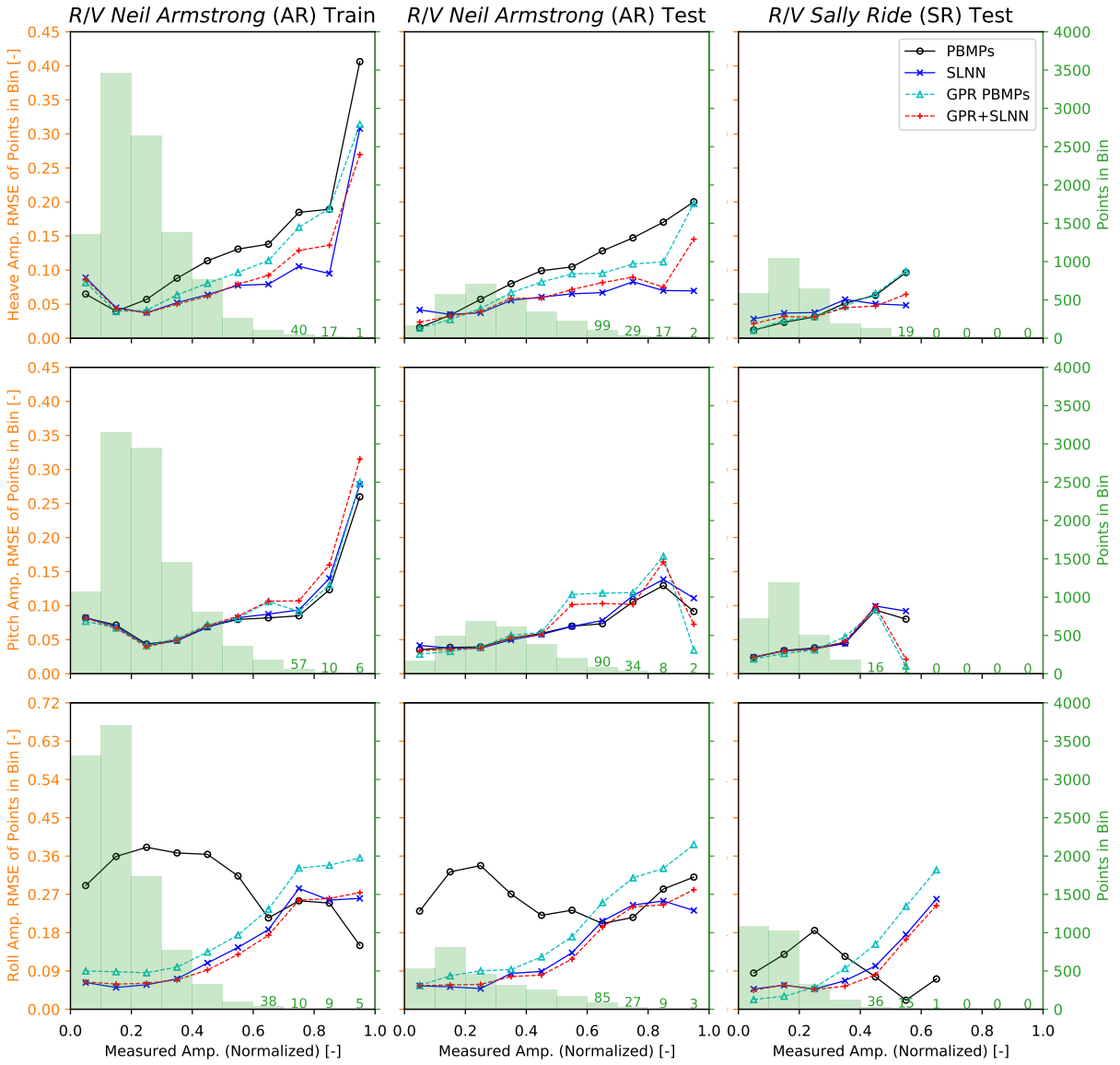

Figure 8.12: Plots of heave, pitch, and roll RMSE as a function of their respective normalized motion amplitude measurements for the CDS wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

Figures 8.13 and 8.14 show the motion amplitude RMSPE and RMSE values for the NWW3 wave data source corresponding to Figures 6.6 and 6.7. For heave, both PBMP curves were very similar to each other for AR Test and SR Test. The GPR PBMPs may have had a slight advantage for AR Test heave overall, but not across all bins. For SR Test pitch, both PBMP curves were very similar, excluding the largest amplitude bin with just one sample. For AR Test pitch, similar performances were seen between PBMPs in intermediate amplitude bins, but the GPR PBMPs were worse at higher amplitudes, which is consistent with the results for CMEMS and CDS. For AR Test roll, the GPR PBMPs were significantly better than the standard PBMPs in all bins. Excluding the highest amplitude bins, which had very few samples, this superior roll performance held for SR Test.

Comparing the SLNN curves in Figures 8.13 and 8.14, they were very similar in most instances across the DOFs. There were select instances where the SLNN outperformed the GPR+SLNN approach, so there is no evidence that the GPR PBMPs assisted the SLNN model structure in any way. Once again, the typical superior performance seen for the SLNN models over the underlying PBMPs demonstrated the predictive power of the SLNN structure.

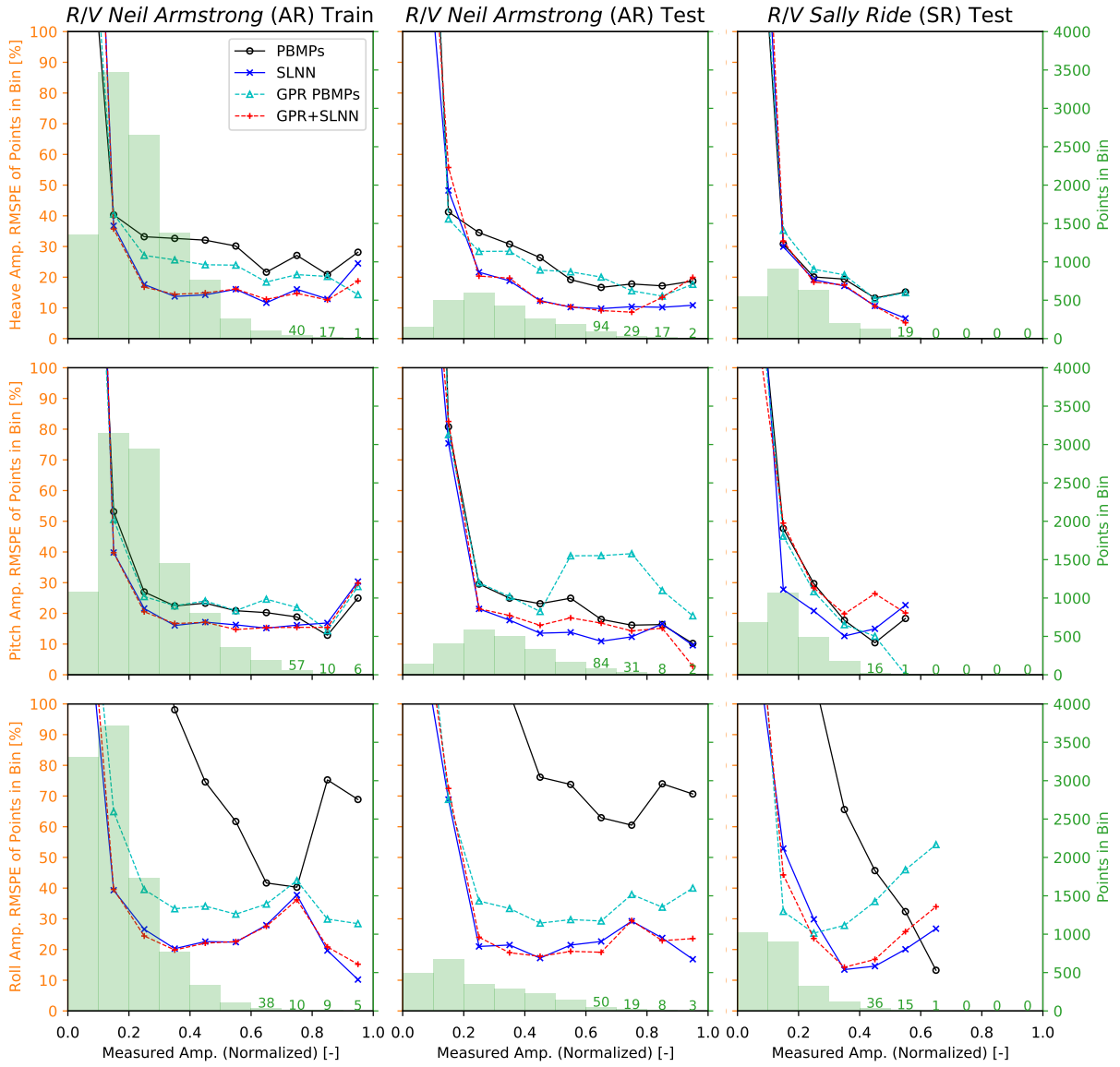

Figure 8.13: Plots of heave, pitch, and roll RMSPE as a function of their respective normalized motion amplitude measurements for the NWW3 wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSPE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

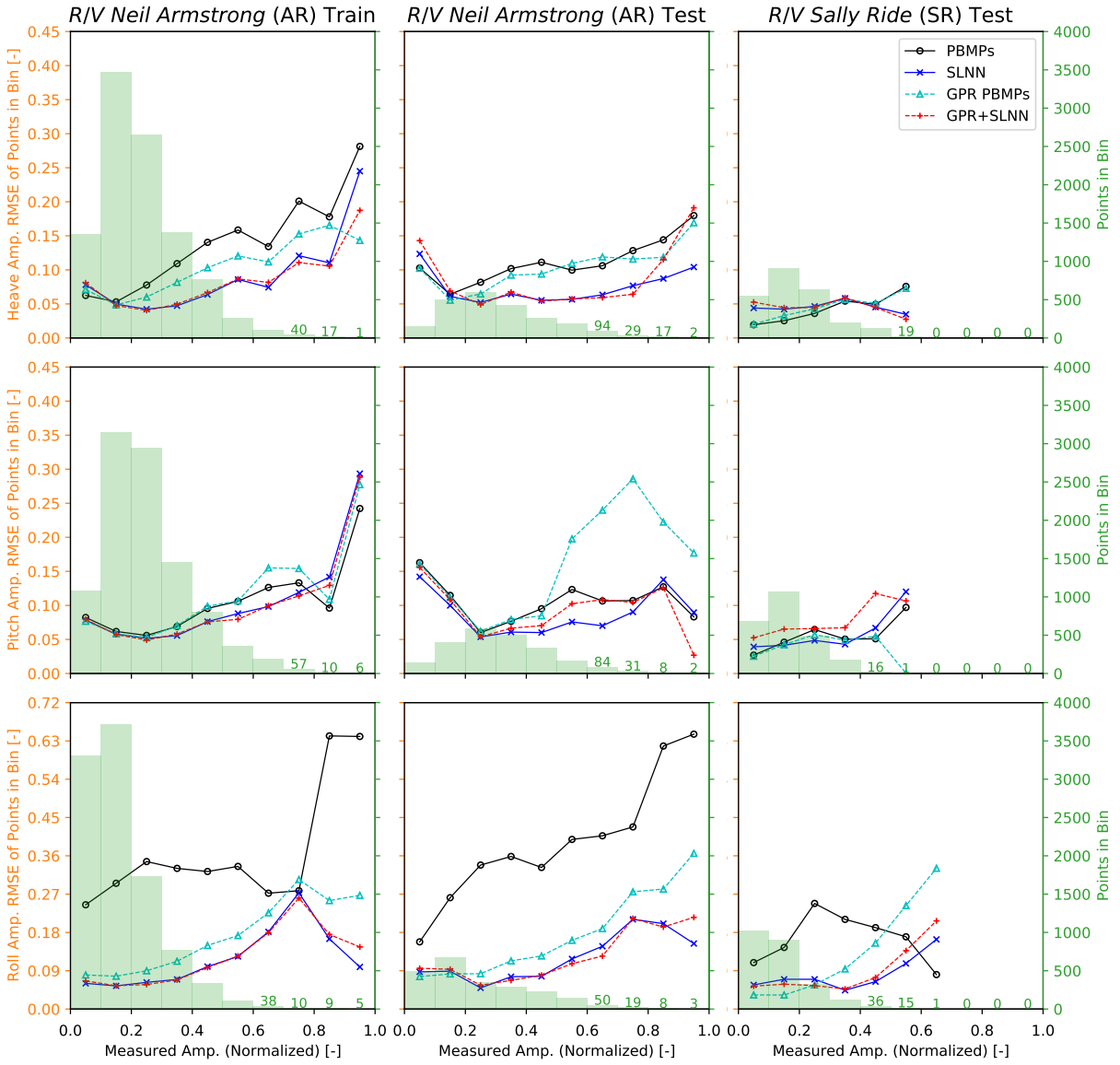

Figure 8.14: Plots of heave, pitch, and roll RMSE as a function of their respective normalized motion amplitude measurements for the NWW3 wave data. Each plot shows a histogram of the number of samples in 10 different measured motion bins, identified by the right axis. The RMSE values, identified by the left axis, calculated from the samples in each bin for a given model or PBMPs, are overlaid on these histograms.

8.3 Full-Scale RAO Correction and Two-Stage Model Demonstration Conclusions

This study developed a novel GPR RAO correction and filtering approach for RAO selection using full-scale motion measurements and 2D spectra from a global wave model. This RAO correction approach represented an update to the underlying physics-based model for improved response prediction capabilities, which addressed the fourth research objective of Chapter I, without requiring data from costly model-scale experiments. The purpose of the filtering approach was to select operational windows and wave frequencies where the vessel's encountered response spectrum could be traced to a specific RAO in wave frequency space with relatively high confidence. The filtering approach considered wave data at the frequency level to avoid discarding entire wave spectra due solely to directional spreading at a few frequencies. Using the resultant measured RAO samples, a GPR correction model was trained, and the GPR RAOs were used in the initial twin framework to generate PBMPS for the three wave data sources and all operational windows considered in Chapter VI. The trained GPR model successfully leveraged GPR's zero-mean property such that RAOs in uninformed regions of the input space reverted to the underlying RAOs. Through comparison to the PBMPs of Chapter VI, it was determined that the GPR correction approach was effective in many instances across the DOFs and wave data sources, but there were select instances in which the standard PBMPs performed better, particularly at greater motion amplitudes. It is possible that these instances were due to the inherent linear assumptions required for the RAO approach, especially for roll, which is highly nonlinear. Nevertheless, it is clear that there are potential improvements for this full-scale RAO correction approach, as outlined in the following paragraph.

First, as noted with respect to the corrected roll RAOs, although instances were

relatively benign and infrequent, it is recommended that any negative RAO values predicted by the GPR model be set to zero to avoid this non-physical scenario. Next, while the use of local maxima in the wave spectra as candidates was logical to maintain a tractable number of samples for GPR and because these maxima corresponded to energy peaks ideal for RAO calculation, this approach may have omitted potential train data samples that would benefit the model. In future studies, it may be worth passing all CDS wave frequencies for each window through the filters, and developing a procedure to down-select the resultant samples to maximize the spread and variance of GPR train data across the input space. Alternatively, sparse Gaussian process estimations, such as those discussed in *Bauer et al.* (2016), have been developed that address the computational cost of GPR training with a large number of points.

The developed wave data filters generated many RAO samples in stern-quartering and following seas conditions that were not usable for RAO correction at first glance; however, there were still notable gaps in the GPR train data at higher speeds. While the consideration of more wave frequencies mentioned in the previous paragraph may mitigate this issue, there are other steps that could address this gap. As mentioned in Chapter VII, a GPR model that considers both full- and model-scale data may be the best solution. For example, stern-quartering and following seas model tests in regular waves, meaning only one wave frequency is present, could directly address gaps in the full-scale data. The filtering approach and GPR model formulation developed in this chapter may also prove useful for irregular wave model test data, such as those considered in Chapter VII.

The second stage of this demonstration, which involved use of the GPR PBMPs to train an SLNN for each wave data source of Chapter VI, did not show any advantage over the SLNNs of Chapter VI. However, the proposed improvements to the RAO correction approach may change this fact in the future. Perhaps the most prominent takeaway from this second demonstration stage matched that of Chapter VII: the

SLNN structure is very effective for improved motion prediction where traditional NN structures would not be suitable. That said, there were many instances where the GPR PBMPs made significant improvements upon the standard PBMPs, showing promise for future implementations. Furthermore, the GPR correction approach is not a black box model like neural network structures (e.g., SLNN), which may be preferred for some implementations.

CHAPTER IX

Conclusions and Future Work

This work explored fusion of global wave model data, physics-based response prediction tools, and onboard measurements using data-driven models for improved vessel response predictions, as demonstrated using over 16,000 30-minute windows from two operational sister ships. In Chapter II, an initial, forward-only twin framework, capable of frequency-domain motion and structural response predictions using wave hindcast or forecast data, as well as fatigue damage tracking, was developed. Not only was this initial twin used to generate PBMPs as data-driven model input variables in subsequent chapters, but it was also demonstrated for fatigue damage tracking and comparisons between four routes in the Pacific Ocean, which showed significantly more damage accumulated on one of the routes. If this trend continued over multiple missions, deployments could be adjusted such that other vessels in the fleet also accumulated similar damage, balancing fleet-wide vessel health using wave data alone. Of course, when available, onboard VBM measurements would allow further improvement in predictive capabilities and fleet damage balancing using data-driven models, as demonstrated in later chapters and highlighted in the following sections. These data-driven approaches represent reflective twin capabilities, which would furnish vessel owners and operators with valuable guidance for underway, deployment, and maintenance decisions.

9.1 Significance of Wave Data Source Selection for Response Prediction and Fatigue Damage Tracking

Given the significant reliance on time-and-place specific wave data for PBMP generation in this work, it was necessary to understand the importance of one's wave data source selection for the initial twin framework. As discussed in Chapter III, wave data and the resultant PBMPs and fatigue damage generated using three different wave data sources were compared at three locations in the Pacific ocean for the entirety of 2017 at 32 speed and heading combinations. In this study, differences between wave data from different sources propagated to the initial twin framework's response predictions, which were then magnified in the calculation of fatigue damage, a higher-order function of stress. These fatigue damage differences were notable after just one year, demonstrating that despite the advanced state of oceanographic models, subtle variations can still cause large discrepancies in the resultant response predictions and fatigue damage estimates. As such, use of higher-fidelity hydro-structural models may not be beneficial in many cases because the quality of results would be limited by uncertainty in the wave data selected.

9.2 The Importance of Retaining Physics-Based Information in Data-Driven Models

Armed with the initial twin framework for PBMP generation and the insights of the wave data source selection study, experiments with linear LS and NN models for motion and VBM prediction using unidirectional wave data were conducted, as discussed in Chapter IV. The successes of these relatively simplistic models using datasets with significant limitations spurred the development of more intensive linear and NN models based on multidirectional wave data in Chapter V. These RR and NN models were trained and tested using over 16,000 30-minute windows from two

operational sister ships. The inclusion of a sister vessel not used in model training demonstrated the potential for sharing data-driven models between ships in a fleet. The vast pool of data allowed construction and comparison of models with and without PBMPs included as input variables, which showed that retaining physics-based information improved model performance, even given over 10,000 training samples. Additionally, this study found that a lack of PBMPs can cause data-driven models to perform more poorly than these PBMPs in some cases, further emphasizing the importance of retaining physics-based information in data-driven models, even given an abundance of train data.

9.3 The Physics-Informed SLNN Structure

Given the clear importance of retaining physics-based information in data-driven models demonstrated in Chapter V, as well as the need for a flexible model structure adaptable to multidirectional wave data with a varying number of directional wave partitions, a novel, physics-informed structure was developed in Chapter VI. The SLNN structure shared layers between directional wave partitions to enforce consistent physics between partitions and allow variation in the number of partitions considered between training samples. This versatility was demonstrated using three unique wave data sources ranging from 3 to 24 maximum wave partitions. While the NN structure from Chapter V was successful given three wave partitions, this structure would quickly become unfeasible as the number of wave partitions increased, especially given variation in this number between samples. The models constructed in Chapter VI also implemented a robust wave data interpolation approach that was immune to changes in a wave model's identification of different wave partitions in space and time, a distinct advantage over the Chapter V models.

9.4 GPR RAO Correction using Full- and Model-Scale Data

Chapter VII targeted seakeeping model inadequacies more directly through GPR correction of RAOs using data from model-scale experiments. While this study demonstrated a potential approach to GPR RAO correction, limitations on the speed and heading combinations available from the model-scale experiments and corresponding full-scale data presented challenges. As such, in Chapter VIII, a GPR RAO correction approach was developed for full-scale data using the large dataset of Chapter V. This novel approach, which avoided the need for expensive model-scale tests, involved a rigorous set of wave data filters to allow calculation of RAOs from full-scale data that were assignable to a specific wave frequency, even in stern-quartering and following seas. The developed method also leveraged GPR's zero-mean property such that RAOs in regions of the input space far from train data samples reverted to the underlying RAOs of the physics-based model. This designed behavior is another example from this work that combats the stigma which suggests data-driven models abandon physics.

A final, two-stage model demonstration was also performed in Chapter VIII that used GPR-corrected RAOs to generate PBMPs (stage one), and these PBMPs were then used to train and test SLNNs (stage two). While this demonstration did not show a distinct benefit to combining the GPR RAO approach with the SLNN structure, it did further demonstrate the power of the SLNN structure to improve upon the underlying PBMPs. Furthermore, there were several instances in which the GPR PBMPs outperformed the standard PBMPs, showing promise for the GPR RAO correction approach. Nevertheless, there are suggested improvements to the GPR RAO correction approach, revisited in the following section, that may make a stronger case for combining the SLNN and GPR models in the future.

9.5 Recommendations for Future Work

Further experimentation with the *Neil Armstrong* and *Sally Ride* dataset is recommended. For example, NN and SLNN models could be retrained using data from both vessels to see if there is any performance improvement for *Sally Ride* test data or any change for the *Neil Armstrong* test data. This demonstration would further examine the feasibility of fleet-wide data-driven models. Furthermore, given the large pool of data available for these ships and their ongoing operation, investigation of time-domain response prediction models is highly recommended.

Given the success of the SLNN model structure demonstrated in this work, experimentation with different input variables is recommended. For instance, given that a corresponding relative wave direction is retained as an input variable for each wave partition, the SLNN structure may prove powerful using the associated 1D wave spectral densities as input rather than period and wave elevation variance metrics. Likewise, as long as speed and relative wave direction variables are retained, the global wave system variables could be replaced with 1D spectral densities. In an even greater extension of this approach, the PBMP input variables and/or the output of the model could be the encountered heave, pitch, and roll response spectral densities rather than just motion variances. Although these variable changes would increase the size of the SLNN, the number of variables would still be independent of the number of wave partitions due to the structure's versatility.

In a less drastic change to the SLNN model input variables, it is also recommended that wind data, which may be significant for roll, are considered for future models. This recommendation also holds for any time-domain approaches developed in future work. Furthermore, evaluation of these approaches for smaller vessels, where nonlinearities may be far more dominant, is recommended.

As recommended previously for the GPR RAO correction approach of Chapter VIII, wave data filtering using more candidate wave frequencies may prove useful to

improve the spread and variance of the GPR train data. Of course, more samples will require investigation of data down-selection approaches or use of sparse GPR methods. Application of the GPR approach developed for full-scale data to the model-scale data of Chapter VII to formally compare methods is also recommended. Finally, given a set of corresponding full- and model-scale data, GPR RAO correction using data from both is recommended. As discussed in VII, rather than a vast amount of expensive experiments, targeted model-scale tests may prove extremely useful for addressing gaps in full-scale RAO measurements.

BIBLIOGRAPHY

BIBLIOGRAPHY

American Bureau of Shipping (2018), Guide for Fatigue Assessment of Offshore Structures, *April 2003, Updated March 2018*.

Ashe, G., et al. (2009), Naval Ship Design , 17th International Ship and Offshore Structures Congress, *2*(August 2009).

Bales, S. L. (1983), Designing ships to the natural environment, *Naval Engineers Journal*, *95*(2), 31–40, doi:10.1111/j.1559-3584.1983.tb00574.x.

Bauer, M., M. Van Der Wilk, and C. E. Rasmussen (2016), Understanding probabilistic sparse Gaussian Process approximations, in *Advances in Neural Information Processing Systems*.

Beck, R. F., and A. W. Troesch (1990), *Student's Documentation and User's Manual for the Computer Program SHIPMO.BM*, University of Michigan Department of Naval Architecture and Marine Engineering.

Bishop, C. M. (2006), *Pattern Recogniton and Machine Learning*, Springer, New York.

Blomquist, B., et al. (2013), The High Wind Gas Exchange Project (HiWinGS): Cruise Report for R/V Knorr KN213-03, 10 Oct to 14 Nov 2013, *Tech. rep.*, CIRES, University of Colorado.

Bremer, K. S. (2018), Using Neural Networks to Predict the Response of a Floating Structure, Master's thesis, Norwegian University of Science and Technology.

Campos, R. M., and C. Guedes Soares (2016), Comparison and assessment of three wave hindcasts in the North Atlantic Ocean, *Journal of Operational Oceanography*, *9*(1), 26–44.

Chollet, F., et al. (2015), Keras, `https://keras.io`.

Copernicus Climate Change Service [C3S] (2021), ERA5: Fifth Generation of ECMWF Atmospheric Reanalyses of the Global Climate. Available online at: https://cds.climate.copernicus.eu/cdsapp#W/home (accessed January, 2021).

De Masi, G., F. Gaggiotti, R. Bruschi, and M. Venturi (2011), Ship motion prediction by radial basis neural networks, *2011 IEEE Workshop on Hybrid Intelligent Models and Applications*, pp. 28–32, doi:10.1109/HIMA.2011.5953967.

Deng, Y., W. Feng, S. Xu, X. Chen, and B. Wang (2020), A novel approach for motion predictions of a semi-submersible platform with neural network, *Journal of Marine Science and Technology*, doi:10.1007/s00773-020-00759-w.

Det Norske Veritas AS (2014), Fatigue Assessment of Ship Structures, *Tech. Rep. 30.7*.

Doherty, B., and Berni Associates Inc. (2004), *R/V Knorr Seabeam Instrumentation. Tech. drawing.*, Falmouth Engineering.

Duz, B., B. Mak, R. Hageman, and N. Grasso (2019), Real time estimation of local wave characteristics from ship motions using artificial neural networks, in *Practical Design of Ships and Other Floating Structures, Proceedings of the 14th International Symposium (PRADS 2019)*, Springer, Yokohama, Japan.

Earle, M. D., K. E. Steele, and D. W. C. Wang (1999), Use of advanced directional wave spectra analysis methods, *Ocean Engineering*, *26*, 1421–1434.

Egan, M. (2015), 'digital twin' technology changed formula 1 and online ads. planes, trains, and power are next.

Fernandez, E., and L. Aouf (2018), PRODUCT USER MANUAL for GLOBAL Ocean Waves Analysis and Forecasting Product GLOBAL_ANALYSIS_FORECAST_WAV_001_027, *Tech. Rep. 1.0*.

Glaessgen, E., and D. Stargel (2012), The digital twin paradigm for future nasa and u.s. air force vehicles.

Gougoulidis, G. (2008), The utilization of artificial neural networks in marine applications: An overview, *Naval Engineers Journal*, *120*(3), 19–26, doi:10.1111/j.1559-3584.2008.00150.x.

GPy (since 2012), GPy: A gaussian process framework in python, `http://github.com/SheffieldML/GPy`.

Guedes Soares, C., and T. Moan (1991), Model uncertainty in the long-term distribution of wave-induced bending moments for fatigue design of ship structures, *Marine Structures*, *4*(4), 295–315.

Hageman, R., M. L. Schirmann, I. Drummen, M. D. Collette, and K. Stambaugh (2019), Sructural Reliability Assessment for a Monitored USCG Cutter, in *ASNE Technology, Systems & Ships 2019*, Washington, DC.

He, K., X. Zhang, S. Ren, and J. Sun (2015), Delving deep into rectifiers: Surpassing human-level performance on imagenet classification, in *2015 IEEE International Conference on Computer Vision (ICCV)*, pp. 1026–1034, doi:10.1109/ICCV.2015.123.

Hersbach, H., et al. (2020), The ERA5 global reanalysis, *Quarterly Journal of the Royal Meteorological Society*, *146*(730), doi:10.1002/qj.3803.

Himeno, Y. (1981), Prediction of ship roll damping - State of the art, *University of Michigan Department of Naval Architecture and Marine Engineering, (Report)*.

Huang, B.-G., and Z.-J. Zou (2016), Short-term prediction of ship pitching motion based on artificial neural networks, in *Proceedings of the ASME 2016 35th International Conference on Ocean, Offshore and Arctic Engineering*, pp. 1–5, Busan, South Korea.

Hughes, O. F. (1988), *Ship Structural Design: A Rationally- Based, Computer-Aided Optimization Approach.*, The Society of Naval Architects and Marine Engineers.

Hulkkonen, T., T. Manderbacka, and K. Sugimoto (2019), Digital Twin for Monitoring Remaining Fatigue Life of Critical Hull Structures, in *18th Conference on Computer Applications and Information Technology in the Maritime Industries (COMPIT2019)*, Tullamore, Ireland.

Innes, M. (2018), Flux: Elegant machine learning with julia, *Journal of Open Source Software*, doi:10.21105/joss.00602.

Jiang, L., S. Signal, B. Jeffries, B. Earley, K. Junghans, D. Hess, and W. Faller (2020), A hydrodynamic digital twin concept for underwater vehicles, in *33rd Symposium on Naval Hydrodynamics (33rd SNH)*, Osaka, Japan.

Kapusuzoglu, B., and S. Mahadevan (2020), Physics-informed and hybrid machine learning in additive manufacturing: Application to fused filament fabrication, *JOM, The Journal of The Minerals, Metals & Materials Society*, *72*(12), 4695–4705, doi:10.1007/s11837-020-04438-4.

Kawai, T., Y. Kawamura, T. Okada, T. Mitsuyuki, and X. Chen (2021), Sea state estimation using monitoring data by convolutional neural network (CNN), *Journal of Marine Science and Technology*, doi:10.1007/s00773-020-00785-8.

Khan, A., C. Bil, and K. E. Marion (2005), Theory and Application of Artificial Neural Networks for the Real Time Prediction of Ship Motion, in *Knowledge-Based Intelligent Information and Engineering Systems*, edited by R. Khosla, R. J. Howlett, and L. C. Jain, pp. 1064–1069, Springer Berlin Heidelberg, Berlin, Heidelberg.

Li, G., B. Kawan, H. Wang, and H. Zhang (2017), Neural-network-based modelling and analysis for time series prediction of ship motion, *Ship Technology Research*, *64*(1), 30–39, doi:10.1080/09377255.2017.1309786.

Lloyd, A. (1989), *Seakeeping: Ship Behavior in Rough Weather*, Ellis Horwood Ltd.

Magoga, T., S. Aksu, S. Cannon, R. Ojeda, and G. Thomas (2019), Through-life hybrid fatigue assessment of naval ships, *Ships and Offshore Structures*, *14*(7), 664–674, doi:10.1080/17445302.2018.1550900.

Malenica, S., Q. Derbanne, F. X. Sireta, F. Bigot, E. Tiphine, G. De-Hauteclocque, and X. B. Chen (2013), HOMER - Integrated hydro-structure interactions tool for naval and off-shore applications, in *RINA, Royal Institution of Naval Architects - International Conference on Computer Applications in Shipbuilding 2013, ICCAS 2013*, vol. 3, pp. 209–221, Royal Institution of Naval Architects.

MARIN (2019), VALID II Joint Industry Project Final Report.

Minnick, L. M., L. W. Hanyok, H. A. Tomaszek, M. P. Melendez, C. R. Turner, J. T. Park, V. Belenky, and C. C. Bassler (2012), Model-Scale Experiment of the Seakeeping Performance for R/V Melville, Model 5720, *Tech. rep.*, Naval Surface Warfare Center Carderock Division, West Bethesda, Maryland.

Mondoro, A., M. Soliman, and D. M. Frangopol (2016), Prediction of structural response of naval vessels based on available structural health monitoring data, *Ocean Engineering*, *125*, 295–307, doi:10.1016/j.oceaneng.2016.08.012.

Najafi, A., H. Nowruzi, and H. Ghassemi (2018), Performance prediction of hydrofoil-supported catamarans using experiment and ANNs, *Applied Ocean Research*, *75*, 66–84, doi:10.1016/j.apor.2018.02.017.

Nielsen, U. D. (2006), Estimations of on-site directional wave spectra from measured ship responses, *Marine Structures*, *19*(1), 33–69, doi:10.1016/j.marstruc.2006.06.001.

Nielsen, U. D., J. J. Jensen, P. T. Pedersen, and Y. Ito (2011), Onboard monitoring of fatigue damage rates in the hull girder, *Marine Structures*, *24*(2), 182–206, doi:10.1016/j.marstruc.2011.03.003.

Nielsen, U. D., A. H. Brodtkorb, and A. J. Sørensen (2018), A brute-force spectral approach for wave estimation using measured vessel motions, *Marine Structures*, *60*, 101–121, doi:10.1016/j.marstruc.2018.03.011.

Nielsen, U. D., A. H. Brodtkorb, and A. J. Sørensen (2019), Sea state estimation using multiple ships simultaneously as sailing wave buoys, *Applied Ocean Research*, *83*, 65–76, doi:10.1016/j.apor.2018.12.004.

Nielsen, U. D., R. E. Mounet, and A. H. Brodtkorb (2021), Tuning of transfer functions for analysis of waveship interactions, *Marine Structures*, *79*, 103,029, doi:https://doi.org/10.1016/j.marstruc.2021.103029.

NOAA National Centers for Environmental Prediction. (2019), *WAVEWATCH III® Production Hindcast, Multigrid: Feb 2005 to October 2018. Accessed date: 01-11-2019.*

NOAA's National Data Buoy Center Center of Excellence in Marine Technology (2018), National Data Buoy Center.

Pedregosa, F., et al. (2011), Scikit-learn: Machine learning in Python, *Journal of Machine Learning Research*, *12*, 2825–2830.

Raissi, M., P. Perdikaris, and G. E. Karniadakis (2019), Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, *378*, 686–707, doi:10.1016/j.jcp.2018.10.045.

Rasmussen, C. E., and C. K. I. Williams (2006), *Gaussian processes for machine learning.*, vol. 14, 1–219 pp., doi:10.1142/S0129065704001899.

Russell, S., and P. Norvig (2010), *Artificial Intelligence: A Modern Approach*, third ed., Prentice Hall.

Schirmann, M., C. Wincott, J. Gose, and M. Collette (2018a), Student poster: Exploration of algorithms and frameworks for surface platform digital twins, *Naval Engineers Journal*, *130*(3), 66–68.

Schirmann, M. L., M. D. Collette, and J. W. Gose (2018b), Ship Motion and Fatigue Damage Estimation via a Digital Twin, in *Life-Cycle Analysis and Assessment in Civil Engineering: Towards an Integrated Vision*, edited by R. Caspeele, L. Taerwe, and D. M. Frangopol, pp. 2075–2082, Taylor & Francis Group, London, Ghent, Belgium.

Schirmann, M. L., T. Chen, M. D. Collette, and J. W. Gose (2019a), Linking Seakeeping Performance Predictions with Onboard Measurements for Surface Platform Digital Twins, in *Practical Design of Ships and Other Floating Structures - Proceedings of the 14th International Symposium, PRADS 2019*, edited by T. Okada, K. Suzuki, and Y. Kawamura, Springer, Yokohama, Japan.

Schirmann, M. L., M. D. Collette, and J. W. Gose (2019b), Impact of weather source selection on time-and-place specific vessel response predictions, in *Trends in the Analysis and Design of Marine Structures*, edited by Parunov and Guedes Soares, pp. 33–41, Taylor & Francis Group, London.

Schirmann, M. L., M. D. Collette, and J. W. Gose (2020a), Significance of wave data source selection for vessel response prediction and fatigue damage estimation, *Ocean Engineering*, *216, 107610*, doi:https://doi.org/10.1016/j.oceaneng.2020.107610.

Schirmann, M. L., M. D. Collette, and J. W. Gose (2020b), Improved Vessel Motion Predictions using Full-Scale Measurements and Data-Driven Models, in *33rd Symposium on Naval Hydrodynamics, 33rd SNH*, Osaka, Japan.

Schirmann, M. L., M. D. Collette, and J. W. Gose (2021), Data-driven models for vessel motion prediction and the benefits of physics-based information, *Revise and resubmit to: Applied Ocean Research*.

SEA-DISTANCES.ORG (2018), Ports distances.

Stambaugh, K., I. Drummen, C. Cleary, R. Sheinberg, and M. Kaminski (2014), Structural fatigue life assessment and sustainment implications for a new class of US coast guard cutters, in *Transactions - Society of Naval Architects and Marine Engineers*.

Stern, F., et al. (2011), Experience from simman 2008 - the first workshop on verification and validation of ship maneuvering simulation methods.

Stopa, J. E., and K. F. Cheung (2014), Intercomparison of wind and wave data from the ECMWF Reanalysis Interim and the NCEP Climate Forecast System Reanalysis, *Ocean Modelling*, *75*, 65–83.

Subramanian, A., and S. Mahadevan (2019), Bayesian estimation of discrepancy in dynamics model prediction, *Mechanical Systems and Signal Processing*, *123*, 351–368, doi:10.1016/j.ymssp.2019.01.014.

The WAVEWATCH III® Development Group (WW3DG) (2016), User manual and system documentation of WAVEWATCH III® version 5.16, *Tech. Note 329*, NOAA/NWS/NCEP/MMAB, College Park, MD, USA.

The Welding Institute (1976), Welding Institute Research Bulletin, *17*(5).

Thompson, I. (2020), Virtual hull monitoring of a naval vessel using hindcast data and reconstructed 2-D wave spectra, *Marine Structures*, *71*, doi:10.1016/j.marstruc.2020.102730.

Thompson, I. M. (2018), Virtual Hull Monitoring: Continuous Fatigue Assessment Without Additional Instrumentation, *International Journal of Maritime Engineering*, *Vol 160*(A3), A–293–A–297, doi:10.3940/rina.ijme.2018.a3.479tn.

U.S. Navy Office of Information (2018), Auxiliary general purpose oceanographic research vessel - AGOR, `https://www.navy.mil/Resources/Fact-Files/Display-FactFiles/Article/2222992/auxiliary-general-purpose-oceanographic-research-vessel-agor/`.

Virtanen, P., et al. (2020), SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python, *Nature Methods*, *17*, 261–272, doi:10.1038/s41592-019-0686-2.

Welch, P. (1967), The use of fast fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms, *IEEE Transactions on Audio and Electroacoustics*, *15*(2), 70–73.

Weymouth, G. D. (2019), Roll damping predictions using physics-based machine learning.

Weymouth, G. D., and D. K. P. Yue (2013), Physics-Based Learning Models for Ship Hydrodynamics, *Journal of Ship Research*, *57*(1), 1–12, doi:10.5957/JOSR.57.1.120005.

Willcox, K. E. (2019), Predictive data science for physical systems: From model reduction to scientific machine learning.

Wirsching, P. M., and M. C. Light (1980), FATIGUE UNDER WIDE BAND RANDOM STRESSES, *Journal of the Structural Division*, *106*(7), 1593–1607.

Wu, M., C. Stefanakos, and Z. Gao (2020), Multi-step-ahead forecasting of wave conditions based on a physics-based machine learning (pbml) model for marine operations, *Journal of Marine Science and Engineering*, *8*(12), 1–24, doi:10.3390/jmse8120992.

Xing, Z., and L. McCue (2010), Modeling Ship Equations of Roll Motion Using Neural Networks, *Naval Engineers Journal*, *122*(3), 49–60, doi:10.1111/j.1559-3584.2010.00241.x.

Zhu, J., and M. Collette (2017), A Bayesian approach for shipboard lifetime wave load spectrum updating, *Structure and Infrastructure Engineering*, *13*(2), 298–312, doi:10.1080/15732479.2016.1165709.